Finite sample sign-based procedures in linear and non-linear statistical models: with applications to Granger causality analysis

SALEHZADEH-NOBARI, KAVEH

FINITE SAMPLE SIGN-BASED PROCEDURES IN LINEAR AND NON-LINEAR STATISTICAL MODELS: WITH APPLICATIONS TO GRANGER CAUSALITY ANALYSIS

Kaveh Salehzadeh Nobari

A thesis submitted for the degree of

Doctor of Philosophy

June 2020

ABSTRACT

This thesis consists of three essays on hypothesis testing and Granger causality analysis. The two main topics under consideration are: (1) exact point-optimal sign-based inference in linear and non-linear predictive regressions with a financial application; and (2) sign-based measures of causality in the Granger sense with an economics application.

These essays can be regarded as an extension to the sign-based procedures proposed by Dufour and Taamouti (2010a). The distinction is that in our study the predictors are stochastic and the signs may exhibit serial dependence. As a consequence, the task of obtaining *feasible* test statistics and measures of Granger causality is more challenging. Therefore, in each essay we either impose an assumption on the sign process or propose tools with which the entire dependence structure of the signs can feasibly be modeled. The three essays are summarized below.

In the first chapter, we acknowledge that the predictors of stock returns (e.g. dividend-price ratio, earnings-price ratio, etc.) are often persistent, with innovations that are highly correlated with the disturbances of the predictive regressions. This generally leads to invalid inference using the conventional T-test. Therefore, we propose point-optimal sign-based tests in the context of linear and non-linear models that are valid in the presence of stochastic regressors. In order to obtain feasible test statistics, we impose an assumption on the dependence structure of the signs; namely, we assume that the signs follow a finite order Markov process. The proposed tests are exact, distribution-free, and robust against heteroskedasticity of unknown form. Furthermore, they may be inverted to build confidence regions for the parameters of the regression function. Point-optimal sign-based tests depend on the alternative hypothesis, which in practice is unknown. Therefore, a problem exists: that of finding an alternative which maximizes the power. To choose the alternative, we adopt the adaptive approach based on the split-sample technique suggested by Dufour and Taamouti (2010a). We present a Monte Carlo study to assess the performance of the proposed "quasi"-point-optimal sign test by comparing its size and power to those of certain existing tests that are intended to be robust against heteroskedasticity. The results show that our procedures outperform the other tests. Finally, we consider an empirical application to illustrate the usefulness of the proposed tests for testing the predictability of stock returns.

In the second chapter, we relax the assumption imposed earlier on the dependence structure of the signs. We had provided a caveat that to obtain feasible test statistics, the Markovian assumption must be imposed on the signs. In this essay, we extend the flexibility of the exact point-optimal sign-based tests proposed in the first chapter, by considering the entire dependence structure of the signs and building feasible test statistics based on pair copula constructions of the sign process. In a Monte Carlo study, we compare the performance of the proposed "quasi"-point-optimal sign tests based on pair copula constructions by comparing its size and power to those of certain existing tests that are intended to be robust against heteroskedasticity. The simulation results maintain the superiority of our procedures to existing popular tests.

In the third chapter, we propose sign-based measures of Granger causality based on the Kullback-Leibler distance that quantify the degree of causalities. Furthermore, we show that by using bound-type procedures, Granger non-causality tests between random variables can be developed as a byproduct of the sign-based measures. The tests are exact, distribution-free and robust against heteroskedasticity of unknown form. Additionally, as in the first chapter, we impose a Markovian assumption on the sign process to obtain feasible measures and tests of causality. To estimate the sign-based measures, we suggest the use of vector autoregressive sieve bootstrap to reduce the bias and obtain bias-corrected estimators. Furthermore, we discuss the validity of the bootstrap technique. A Monte Carlo simulation study reveals that the bootstrap bias-corrected estimator of the causality measures produce the desired outcome. Furthermore, the tests of Granger non-causality based on the signs perform well in terms of size control and power. Finally, an empirical application is considered to illustrate the practical relevance of the sign-based causality measures and tests.

Keywords: Stochastic regressors; stock return predictability; valuation ratios; persistency; sign test; point-optimal test; non-linear model; heteroskedasticity; exact inference; distribution-free; split-sample; adaptive method; projection technique; numerical optimization; causality measures; time series; Kullback-Leibler distance; bootstrap; Bonferroni test; D-vine; power envelope.

To my parents, sister and brother.

Acknowledgments

First and foremost, I wish to express my gratitude to my supervisor, Professor. Abderrahim Taamouti, for his continuous support, patience and encouragement throughout the Ph.D. process. Completing this thesis would have been an impossible task without his vast knowledge on the subject area and his research experience. Professor. Taamouti has been a great source of inspiration and often sacrificed his valuable personal time (including weekends and public holidays) to ensure that I am making satisfactory progress on the thesis and related projects.

Besides my supervisor, I also wish to thank Dr. Henk Slim who was responsible for maintaining the High Performance Computing (HPC) cluster at Durham. My research was heavily simulations intensive, which implied that I was a constant nuisance to Dr. Slim. He was incredibly helpful and patient and always made sure that any issues that I encountered with the cluster were resolved in a timely manner.

I will forever be indebted to my parents, as well as my sister and my brother, without whom none of this would have been possible. They supported me through thick and thin, not just in the duration of writing this thesis, but in life in general. I am grateful to them for always encouraging me in difficult times and for going out of their way to make sure that I can comfortably follow through with the plans that I have laid out for myself.

Last but not least, I would also like to thank the support provided by my friends and relatives. In no particular order, special thanks goes to my brother's wife, Sara, for her kind and hospitable nature, my cousins, Dr. Elnaz Salehzadeh Nobari and (soon-to-be) Dr. Amin Salehzadeh Nobari, as well as my dear friends, Dr. Ruslan Tuneshev, Ms. Helena Brennan, Mr. Weidong Lin, Dr. Magomet Aushev, Mr. Gianluca Guardo and Mr. Ernesto Saputo for their constant help, encouragement and understanding, particularly for their patience during what appeared to be a perpetual state of stress and anxiety in the final months of my Ph.D. I would also like to thank my uncle, Dr. Ali Salehzadeh Nobari, as well as my friends, Mr. Arash Yari, Mr. Gonzalo Avila Lafuente, Ms. Charlotte Chopping, Dr. Jiunn Wang, and Dr. XiaoXiao Ma for always offering moral support. Individuals in this list have all been instrumental in the completion of this thesis; however, I must note that this list is by no means exhaustive.

Contents

List of Tables

List of Figures

Motivation

It is well established that in nonparametric settings and under weak distributional assumptions, it is not possible to obtain a valid test or confidence interval on the mean (moments) of observations, even in large sample sizes and under the restrictive assumption of i.i.d data. More specifically, the absence of the knowledge of the population distribution implies that a sample from the population would provide little information about its tails; and as the population mean μ is sensitive to the tails of the distribution, there are no effective means of testing the hypotheses of the form, say, $H_0 : \mu = 0$ [see Bahadur and Savage (1956)]. These findings suggest that a valid test (i.e. control size whatever the sample size) about the mean μ with level α, where $0 \leq \alpha < 1$, can not possess power greater than its size when the distribution is unspecified. This situation which is referred to as the issue of non-testability extends to the coefficients of regression models [see Dufour et al. (2008)]. Furthermore, the simulation results of Dufour and Taamouti (2010a), as well as in the first and the second chapters of this thesis, reveal that in the presence of small number of observations and under different forms of heteroskedasticity, "robust"test statistics, such as the T-test based on White (1980) variance-correction suffer from poor power and size control. These issues have turned the attention of numerous scholars to quantiles (e.g. median). Under adequately broad distributional assumptions, sign statistics are the only possible way of producing valid inference about the median in finite sample procedures and in the presence of general forms of heteroskedasticity [see Lehmann et al. (1949) and Pratt and Gibbons (2012)]. This is due to the fact that sign statistics are predicated on quantiles, which do not suffer from the issue of non-testability. As such, they offer more flexibility, since no moment conditions are imposed on the dependent variable. Motivated by the generalities of sign statistics, this thesis proposes sign-based procedures in the context of linear and non-linear models that are robust against non-standard and asymmetric distributions and further address certain gaps within the domains of nonparametric inference for time-series data.

To further elaborate on some of the points raised earlier, consider testing the hypothesis that a sample of n observations are independent with a distribution symmetric about zero

$$H_0 : X_1, \cdots, X_n \quad \text{are independent with a common zero median.}$$

Rejecting this null hypothesis would imply that either the distribution is asymmetric, or that it has a median other than zero. Further, If the assumption of symmetry is imposed, this would turn to a location hypothesis, as $H_0 : \mu = 0$. Given the absolute values of the observations $|X_1|, \cdots, |X_n|$, under the null hypothesis the assignments of positive and negative signs are equally likely, and since there are 2^n possible permutations of assigning the signs, there are 2^n members of the family of possible samples, each with a probability of 0.5^n. Hence, once the test criterion is selected, the test statistic can be calculated for each member of the family and the distribution of the test statistic is obtained under the null hypothesis, using which the critical values are easily found. Pratt and Gibbons (2012) refer to this testing approach and its respective simulated distribution under the null hypothesis, as the randomization test and the randomization distribution of the test statistic respectively, which are conditional on the absolute values of the observations. The level of randomization test is α if

$$P\left[\text{Rejecting } H_0 \mid |X_1|, \cdots, |X_n|\right] \leq \alpha$$

under the null hypothesis. In other words, this implies that any valid test must have level equal to α conditional on the absolute values, otherwise the procedure has size one. It is not immediately obvious whether the robust least-squares based T-test based on White (1980) variance correction (commonly designated as "HAC") satisfy the above condition.

The literature surrounding sign-based inference is vast with numerous scholars dedicating articles, books and monographs to study it extensively. Sign-based procedures started to attract careful attention after decades of research on rank-based inference in the area of nonparametric statistics. The latter concerns the rank of the observations (specifically that of the residuals of the fitted data), as opposed to their numerical value. On the other hand, experiments in sign-based inference in the past few decades mainly express interest on the signs of the observations. As with the example above, under the general assumption that the random residuals possess positive and negative signs with equal probability, or in other words

$$P[\varepsilon_t > 0] = P[\varepsilon_t < 0], \quad t = 1, \cdots, n$$

sign-based tests are distribution-free.

Boldin et al. (1997) develop sign-based procedures in the context of independent and time-series data for linear statistical models. They particularly focus on exact inference (i.e. results that provide exact significance level) in finite samples and further explore the asymptotic properties of sign-based statistics. In a more recent work, Taamouti (2015) surveys the latest developments in sign-based inference. These include procedures for testing orthogonality between random variables in linear and non-linear models and in the context of both independent and dependent data. By considering a general regression of the form

$$y_t = \mu + f(x_t; \beta) + \varepsilon_t,$$

Taamouti (2015) reviews an array of sign tests that impose different assumptions on the residuals ε_t, the functional form $f(x_t; \beta)$, the randomness and the dimension of x_t, and on the presence or the absence of a drift parameter μ For independent data, many studies propose distribution-free sign and sign-ranked statistics that are exact and robust against different forms of heteroskedasticity. A few notable examples in the context of one regressor (where x_t is a scalar) include the sign-based procedures of Campbell and Dufour (1991, 1995, 1997) and Luger (2003) among others, where the said procedures are shown to be exact in the presence of non-Gaussian, asymmetric and heteroskedastic distributions. The sign-based and Wilcoxon sign-ranked test statistics proposed by Campbell and Dufour (1991, 1995) are non-parametric analogues of the T-test in the absence of a nuisance drift parameter μ, whereas Campbell and Dufour (1997) and Luger (2003) extend these procedures to further incorporate a drift term. In a multivariate framework within the context of independent data, where x_t is a $k \times 1$ vector of fixed regressors, Dufour and Taamouti (2010a) propose exact tests that further address the issue of optimality (i.e. maximum achievable power for a given testing problem), under the broad assumption that the residuals possess zero median conditional on the explanatory variables:

$$P[\varepsilon_t > 0 \mid X] = P[\varepsilon_t < 0 \mid X] = \frac{1}{2}, \quad t = 1, \cdots, n,$$

where $X = [1, x_1, \cdots, x_n]'$ is an $n \times (k+1)$ matrix of explanatory variables. On the other hand, in the context of dependent data, Coudin and Dufour (2009) develop sign-based statistics that further consider serial (non-linear) dependence and discrete distributions. To construct their tests, they impose a mediangale assumption on the disturbances, which is a median-based analogue of a martingale difference sequence. The process of disturbances is a weak conditional mediangale iff

$$P[\varepsilon_1 > 0 \mid X] = P[\varepsilon_1 < 0 \mid X], \quad \text{and} \quad P[\varepsilon_t > 0 \mid \varepsilon_1, \cdots, \varepsilon_{t-1}, X] = P[\varepsilon_t < 0 \mid \varepsilon_1, \cdots, \varepsilon_{t-1}, X],$$

for $t = 2, \cdots, n$. This assumption permits ε_t to possess discrete distributions, since it allows for non-zero probability mass at zero. A strict version of the conditional mediangale is

$$P[\varepsilon_1 > 0 \mid X] = P[\varepsilon_1 < 0 \mid X] = \frac{1}{2},$$

and

$$P[\varepsilon_t > 0 \mid \varepsilon_1, \cdots, \varepsilon_{t-1}, X] = P[\varepsilon_t < 0 \mid \varepsilon_1, \cdots, \varepsilon_{t-1}, X] = \frac{1}{2}, \quad \text{for} \quad t = 2, \cdots, n$$

which has no mass at zero. For the purpose of simplifying the construction of their test statistic, Coudin and Dufour (2009) impose the latter assumption.

In this thesis, we extend the work of Dufour and Taamouti (2010a), by constructing sign-based procedures for linear and non-linear predictive regressions of the form

$$y_t = f(x_{t-1}; \beta) + \varepsilon_t,$$

where the data is no longer necessarily independent and x_{t-1} is a $(k+1) \times 1$ vector of fixed or stochastic regressors. In the first and the second chapters, we propose exact point-optimal (i.e.

optimal at a specific point in the alternative hypothesis parameter space) sign-based inference to test for orthogonality between random variables. These tests are motivated by the type of feedback studied by Mankiw and Shapiro (1986), in which x_t is assumed to follow an AR(1) process

$$x_t = \theta x_{t-1} + u_t$$

with

$$cov(\varepsilon_t, u_t) = \rho \quad \text{and} \quad corr(u_{t+j}, \varepsilon_t) = 0, \quad \forall j \neq 0.$$

Thus, while x_{t-1} is contemporaneously uncorrelated with ε_t[1], it is not uncorrelated with all its leads and lags. Particularly x_{t-1} is correlated with $\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots$ and as such $\mathbb{E}[\varepsilon_t \mid x_t, x_{t-1}] \neq 0$, which is a violation of the Gauss-Markov theorem. As a consequence of this, finite-sample estimation and inference are misleading, as the OLS estimate of β is biased and has a non-standard sampling distribution [see Stambaugh (1999) for a practical example]. Using Monte-Carlo simulations, Mankiw and Shapiro (1986) examine the actual rejection rates in T-type tests of the hypothesis $H_0 : \beta = 0$ and find that when $\theta < 1$ yet the process x_t is highly persistent, the test does not control size in finite samples; albeit, the size distortions improve as $n \to \infty$. However, when $\theta = 1$, the size distortions persist as $n \to \infty$. As noted by Magdalinos and Phillips (2009), when the process x_t is mildly integrated towards the stationary side (with $\theta < 1$), the OLS estimator is consistent and asymptotically normal, yet suffers from significant bias in finite samples. On the other hand, in the moderately explosive case (with $\theta > 1$), the least squares estimator is mixed normal with Cauchy-type tail behavior with an explosive convergence rate.

By imposing the strict conditional mediangale assumption on the residuals, which allows for serial (non-linear) dependence between $y_1, \cdots, y_n$, the point-optimal sign-based tests (POS-based tests hereafter) introduced in this thesis are distribution-free and exact in the presence of non-standard distributions - for instance, in the context of predictive regressions with highly persistent and potentially non-stationary regressors. Furthermore, they are robust against heteroskedasticity of unknown form and "trace out"the power envelope [see King (1987)], and can be inverted to

[1]Note that as highlighted by Banerjee et al. (1993) this is not a simultaneity problem - i.e. $corr(x_{t-1}, \varepsilon_t) = 0$

produce confidence region for the vector (sub-vector) of parameters.

POS-based tests depend on the alternative hypothesis, which in practice is unknown. Therefore, a problems consists of finding an alternative such that the power curve of the POS-based test is close to that of the power envelope. In the first chapter, we follow Dufour and Taamouti (2010a) by using 10% of the sample to find the alternative and the rest to calculate the test statistic to maximize the power of the POS-based tests.

In the second chapter, we relax the Markovianity assumption by considering the entire dependence structure of the signs using copulae. However, as the signs are discrete by nature, evaluating the sign statistics with a sample size n involves 2^n *multivariate* copula evaluations, which is computationally infeasible. Therefore, we propose exact point-optimal sign-based tests based on pair copula constructions (PCC hereafter) of discrete data using models and algorithms introduced by Panagiotelis et al. (2012), which drastically reduce the computations to $2n(n-1)$ *bivariate* copula evaluations. Furthermore, by conducting an extensive simulations exercise in the second chapter, we extend the optimal split-sample ratio proposed in the first chapter to the POS-based tests constructed using the pair copula constructions of the signs. Our simulation study reveals that the point-optimal sign-based tests based on PCC using the 10% sample-split outperform the existing popular tests in the majority of circumstances.

Finally, in the third chapter, we extend the work of Gourieroux et al. (1987) and propose sign-based Granger causality measures based on the Kullback-Leibler distance criterion. The proposed measures assess the strength of the relationship between random variables and quantify the degree of causalities. Furthermore, we show that by using the bound-type procedures suggested by Dufour (1990) and Campbell and Dufour (1997) to address the nuisance parameter problem, Granger non-causality tests can be developed as a byproduct of the sign-based Granger causality measures. These tests are exact, distribution-free and robust against heteroskedasticity.

To obtain feasible sign statistics and measures of Granger causality in the first and the third chapters, we first impose an assumption on the dependence structure of the sign processes; specifically, we assume that the signs follow a Markov process of order one. The simulations reveal that in spite of the Markovianity assumption, the sign-based tests of orthogonality outperform certain existing tests that are intended to be robust against non-standard distributions and heteroskedasticity of unknown form. Furthermore, the sign-based Granger non-causality tests control size and possess good power properties.

Chapter 1

Exact point-optimal sign-based tests for predictive linear and non-linear regressions

1.1 Introduction

Numerous studies investigate the predictability of financial and economic variables using the past values of one or more predictors. The most commonly encountered examples of such studies concern the predictability of stock returns using the lag of certain fundamental variables, such as the dividend-yield, earnings-price ratio or interest rates [see Campbell and Shiller (1988), Fama and French (1988), Campbell and Yogo (2006), Campbell and Thompson (2008), and Golez and Koudijs (2018), among others]. Predictability in this context is generally assessed using the OLS regression of the returns against the said predictors and tested with conventional T-type tests. However, the predictors that are often considered in these studies are known to be highly persistent with innovations that are correlated with the disturbances in the predictive regression of the returns. In such situations, we know that the OLS estimator of the coefficients, although consistent, will be biased. As a result of this bias, in finite samples the T-statistic will have a nonstandard distribution which leads to invalid inference [see Mankiw and Shapiro (1986), Banerjee et al.

(1993) and Stambaugh (1999) among others]. In this paper, we address this issue by deriving point-optimal sign-based tests (hereafter POS-based tests) in the context of linear and non-linear predictive regressions that are distribution-free, robust against heteroskedasticity of unknown form and which allow for serial (non-linear) dependence provided that the residual process has zero median conditional on the explanatory variables and its own past. This assumption allows the signs to be i.i.d under the null hypothesis of orthogonality according to a known distribution, despite the fact that the variables to which the indicator functions are applied are dependent [see Coudin and Dufour (2009)].

Nelson and Kim (1993) reduce the small-sample bias using bootstrap simulations and Stambaugh (1999) shows that in the case of stationary regressors the said bias can be corrected. However, in later studies Phillips and Lee (2013) and Phillips (2014) show this to be infeasible in the presence of predictors that exhibit local-to-unity, unit-root or explosive persistency. Therefore, many inference procedures in this context address the issue of size distortions by considering local-to-unity asymptotics, where the key predictor variable is assumed to contain a unit root [Lewellen (2004)], or can be modeled as having a local-to-unit root [Elliott and Stock (1994), Torous et al. (2004), and Campbell and Yogo (2006), among others]. Notable studies under the local-to-unity dynamics employ an array of procedures, such as Bonferroni corrections [e.g. Cavanagh et al. (1995) and Campbell and Yogo (2006)], a conditional likelihood based approach [e.g. Jansson and Moreira (2006)], as well as the nearly optimal tests proposed by Elliott et al. (2015). In more recent work, Kostakis et al. (2015) and Phillips and Lee (2016) expand on the predictability literature by utilizing an extension of the instrumental variable procedure suggested by Phillips et al. (2009) to generalize inference to multivariate regressors with stationary, local-to-unity and explosive persistency. The contribution of the POS-based tests proposed in our study is twofold: firstly, as our tests are distribution-free, they are valid in the presence of regressors with general persistency in finite samples and do not suffer from the discontinuity that is commonly observed in the limiting distribution of conventional test statistics between stationary, local-to-unity and explosive autoregressions. Secondly, our tests possess the greatest power among certain parametric and non-parametric tests that are often encountered in practice and can easily be extended to multivariate testing problems.

In a recent study, Dufour and Taamouti (2010a) propose simple point-optimal sign-based tests in the context of linear and non-linear regression models, which are valid under non-normality and heteroskedasticity of unknown form, provided the errors have zero median conditional on the explanatory variables. These tests are exact, distribution-free, and robust against heteroskedasticity of unknown form, and may be inverted to build confidence regions for the vector of unknown parameters. This work, however, is developed under the assumption that the predictors are fixed; thus, these tests are not applicable in the presence of stochastic regressors. We extend the above tests to the case where the predictors can be fixed or stochastic. The main motivation is to build point-optimal sign-based tests for linear and non-linear predictability of stock returns that retain the advantages of the POS-based tests proposed by Dufour and Taamouti (2010a).

To extend the previous work of Dufour and Taamouti (2010a), we recognize that under the alternative hypothesis the signs are no longer necessarily independent and the test-statistic now depends on the joint distribution of the signs, which is computationally infeasible. Therefore, an additional assumption on the dependence structure of the process of signs is needed to obtain *feasible* test statistics. In particular, we assume that this process is a Markov process of finite-order. By construction, our POS-based tests control the size for any given sample. Under the null hypothesis of unpredictability, the tests are valid even in the presence of the bias problem pointed out by Mankiw and Shapiro (1986) and Stambaugh (1985, 1999), which affects the classical testing procedure for stock returns predictability. In addition, our tests are model-free and robust against heteroskedasticity of unknown form. The tests are point-optimal tests, which are useful in a number of ways and are particularly attractive when testing one financial theory against another. An important feature of these tests stems from the fact that they trace out the power envelope, i.e. the maximum achievable power for a given testing problem, which may be used as a benchmark against which other testing procedures can be evaluated. Finally, our tests may be inverted to build confidence regions for the parameters of the regression function.

As point-optimal tests maximize the power at a nominated point in the alternative hypothesis parameter space, a practical problem concerns finding an alternative at which the power curve of the POS-based test is close to that of the power envelope. Following Dufour and Torrès (1998), Dufour and Jasiak (2001) and Dufour and Taamouti (2010a), we propose an adaptive approach based on the split-sample technique to choose the alternative hypothesis. This procedure consists of splitting the sample into two independent sub-samples, where the first part is used to estimate the alternative hypothesis and the second part to compute the POS-based test statistic [see Dufour and Iglesias (2008)]. In a simulations exercise, Dufour and Taamouti (2010a) find that using the first 10% of the sample to estimate the alternative and the rest to compute the test statistic, achieves a power that traces out the power envelope. We present a Monte Carlo study to assess the performance of the proposed "quasi"-POS-based tests by comparing its size and power to certain existing tests that are intended to be robust against heteroskedasticity. We show the superiority of our procedures in the presence of nearly integrated regressors and under different distributional assumptions and forms of heteroskedasticity..

The plan of the paper is as follows: Section 1.2 provides exact POS-based tests in the context of linear and non-linear predictive regressions. Section 1.3 discusses the selection approach for the alternative hypothesis to compute the POS-based test statistic. Section 1.4, discusses the construction of POS-based confidence regions using the projection techniques. Section 1.5, presents a Monte Carlo study to assess the performance of the POS-based tests by comparing their size and power to those of certain popular tests. Section 1.6 is devoted to an empirical application, and finally, the paper is concluded in Section 1.7. Proofs are presented in Appendix 1.8.

1.2 POS tests in linear and non-linear regression models

In this Section, we follow the structure of Dufour and Taamouti (2010a) to derive POS-based tests in the context of linear and non-linear regression models. However, in our study stochastic regressors may as well be considered. First, we divert our attention to the problem of testing whether the conditional median of a vector of observations is zero against a linear regression alternative. This is is later generalized to test whether the coefficients of a possibly non-linear median regression function have a given value against an alternative non-linear median regression. Although the former problem is a special case of the latter, for the simplicity of exposition the linear regression model is considered first.

1.2.1 Testing independence (zero coefficients) hypothesis in linear regressions

Consider a stochastic process $Z = \{Z_t = (y_t, x_t') : \Omega \to \mathbb{R}^{(k+1)} : t = 0, 1, \cdots\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. Let $\{Z_t, \mathcal{F}_t\}_{t=0,1,\cdots}$ be an adapted stochastic sequence, such that $\mathcal{F}_t$ is a σ-field in Ω, $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s < t$, $\sigma(Z_0, \cdots, Z_t) \subset \mathcal{F}_t$, where $\sigma(Z_0, \cdots, Z_t)$ is the σ-field generated by $Z_0, \cdots, Z_t$. Suppose that y_t can linearly be explained by a vector variable x_t

$$y_t = \beta' x_{t-1} + \varepsilon_t, \quad t = 1, ..., n, \tag{1.1}$$

where x_{t-1} is an $(k+1)\times 1$ vector of stochastic explanatory variables, say $x_{t-1} = [1, x_{1,t-1}, ..., x_{k,t-1}]'$, $\beta \in \mathbb{R}^{(k+1)}$ is an unknown vector of parameters with $\beta = [\beta_0, \beta_1, ..., \beta_k]'$ and

$$\varepsilon_t \mid X \sim F_t(. \mid X)$$

where $F_t(. \mid X)$ is an unknown conditional distribution function and X=$[x_0, \cdots, x_{n-1}]'$ is an $n \times (k+1)$ matrix.

In the context of general forms of (non-linear) dependence, an assumption that is commonly made on the error terms $\{\varepsilon_t, t = 1 \cdots, n\}$ is that the error process is a martingale difference sequence (MDS hereafter) with respect to $\mathcal{F}_t = \sigma(Z_0, \cdots, Z_t)$ for $t = 0, 1, \cdots,$ - i.e. $\mathbb{E}\{\varepsilon_t \mid \mathcal{F}_{t-1}\} = 0$, $\forall t \geq 1$. As the latter assumption relies on the first moment of the residuals, we follow Coudin and Dufour (2009) by departing from this assumption and considering the median as an alternative measure of central tendency. This implies imposing a median-based analogue of the MDS on the error process - namely we suppose that ε_t is a strict conditional mediangale

$$P[\varepsilon_t > 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = P[\varepsilon_t < 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = \frac{1}{2}, \tag{1.2}$$

with

$$\boldsymbol{\varepsilon}_0 = \{\emptyset\}, \quad \boldsymbol{\varepsilon}_{t-1} = \{\varepsilon_1, \cdots, \varepsilon_{t-1}\}, \quad \text{for} \quad t \geq 2$$

Note (1.2) entails that $\varepsilon_t \mid X$ has no mass at zero for all t, which is only true if $\varepsilon_t \mid X$ is a continuous variable. Model (1.1) in conjunction with assumption (1.2) allows the error terms to possess asymmetric, heteroskedastic and serially (non-linear) dependent distributions, so long as the conditional medians are zero. Assumption 1.2 allows for many dependent schemes, such as those of the form $\varepsilon_1 = \sigma_1(x_0, \cdots, x_{t-2})\epsilon_1$, $\varepsilon_t = \sigma_t(x_0, \cdots, x_{t-2}, \varepsilon_1, \cdots, \varepsilon_{t-1})\epsilon_t$, $t = 2, \cdots, n$, where $\epsilon_1, \cdots, \epsilon_n$ are independent with a zero median. In time-series context this includes models such as ARCH, GARCH or stochastic volatility with non-Gaussian errors. Furthermore, in the mediangale framework the disturbances need not be second order stationary.

We wish to test the null hypothesis

$$H_0 : \beta = 0 \tag{1.3}$$

against the alternative H_1

$$H_1 : \beta = \beta_1. \tag{1.4}$$

We define the following vector of signs

$$U(n) = (s(y_1), ..., s(y_n))',$$

where, for $1 \leq t \leq n$,

$$s(y_t) = \begin{cases} 1, \; if \;\; y_t \geq 0 \\ 0, \; if \;\; y_t < 0 \end{cases}.$$

The test is Neyman-Pearson type test based on signs [see Lehmann and Romano (2006)] which maximize the power function under the constraint $P\,[\text{reject } H_0 \mid H_0] \leq \alpha$. The idea is to build point-optimal sign-based tests to test the null hypothesis (1.3) against the alternative hypothesis (1.4). To do so, we first define the likelihood function of sample in terms of signs $s(y_1), ..., s(y_n)$

conditional on X

$$L\left(U(n),\beta\right)=P\left[s(y_1)=s_1,...,s(y_n)=s_n\mid X\right]=\prod_{t=1}^{n}P\left[s(y_t)=s_t\mid \underset{\sim}{\mathrm{S}}_{t-1},X\right],$$

with

$$\underset{\sim}{\mathrm{S}}_0=\{\emptyset\}\,,\quad \underset{\sim}{\mathrm{S}}_{t-1}=\{s(y_1)=s_1,...,s(y_{t-1})=s_{t-1}\}\,,\ \text{for } t\geq 2,$$

and

$$P\left[s(y_1)=s_1\mid \underset{\sim}{\mathrm{S}}_0,X\right]=P\left[s(y_1)=s_1\mid X\right],$$

where each s_t, for $1\leq t\leq n$, takes two possible values 0 and 1.

As the error terms satisfy the strict conditional mediangale assumption (1.2), the distribution of the signs $s(\varepsilon_1),\cdots,s(\varepsilon_n)$, and in turn $s(y_1),\cdots,s(y_n)$ under the null hypothesis of orthogonality is specified and are mutually independent [see Coudin and Dufour (2009)].

Theorem 1 *Under model (1.1) and assumption (1.2), the variables $s(\varepsilon_1),\cdots,s(\varepsilon_n)$ are i.i.d conditional on X, according to the distribution*

$$P[s(\varepsilon_1)=1\mid X]=P[s(\varepsilon_1)=0\mid X]=\frac{1}{2},\quad t=1,\cdots,n$$

This result holds true iff for any combination of $t=1,\cdots,n$ there is a permutation $\pi:i\rightarrow j$ such that the mediangale assumption holds for j. Then the signs $s(\varepsilon_1),\cdots,s(\varepsilon_n)$ are i.i.d.

Proof: *See Appendix.*

A sign-based test for testing the null hypothesis (1.3) against the alternative hypothesis (1.4) is given by the following proposition.

Proposition 1 *Under assumptions (1.2) and (1.1), let H_0 and H_1 be defined by (1.3) - (1.4),*

$$SL_n(\beta_1)=\sum_{t=1}^{n}a_t(\beta_1)s(y_t),$$

where, for $t=1,...,n$,

$$a_t(\beta_1)=\ln\left\{\frac{P\left[y_t\geq 0\mid \underset{\sim}{S}_{t-1},X\right]}{P\left[y_t<0\mid \underset{\sim}{S}_{t-1},X\right]}\right\}, \tag{1.5}$$

and suppose the constant $c_1(\beta_1)$ satisfies $P\left[\sum_{t=1}^{n} a_t(\beta_1)s(y_t) > c_1(\beta_1)\right] = \alpha$ under H_0, with $0 < \alpha < 1$. Then the test that rejects H_0 when

$$SL_n(\beta_1) > c_1(\beta_1) \tag{1.6}$$

is most powerful for testing H_0 against H_1 among level-α tests based on the signs $\left(s(y_1), ..., s(y_n)\right)'$.

Proof: *See Appendix.*

Notice that the calculation of the test statistic $SL_n(\beta_1)$ depends on the weights $a_t(\beta_1)$, which in turn depends on the calculation of the conditional probabilities $P\left[y_t \geq 0 \mid \underset{\sim}{S}_{t-1}, X\right]$ and $P\left[y_t < 0 \mid \underset{\sim}{S}_{t-1}, X\right]$. The latter terms are not easy to compute and involves the distribution of the joint process of signs $\left(s(y_1), ..., s(y_n)\right)'$ conditional on X, which is unknown. An alternative way to compute the terms $P\left[y_t \geq 0 \mid \underset{\sim}{S}_{t-1}, X\right]$ and $P\left[y_t < 0 \mid \underset{\sim}{S}_{t-1}, X\right]$ is to use simulations, however again this will be time consuming as it requires to simulate the joint distribution of the process of signs $\left(s(y_1), ..., s(y_n)\right)'$, which depends on the sample size n. For all these reasons and to make the test statistic $SL_n(\beta_1)$ feasible, we make the following assumption.

Assumption A1: *Let $\{y_t, t = 0, 1, \cdots\}$ follow a Markov process of order one. Then under the alternative hypothesis, the sign process $\{s(y_t)\}_{t=0}^{\infty}$ follows a Markov process of the same order.*

Proof: *See Appendix.*

As the mediangale assumption allows for non-linear serial dependence, testing assumption **A1** by considering linear correlation is inappropriate. One approach involves fitting copula models, which provides the means of separating the marginal distributions of the process from their respective dependence structure. The latter stems from Sklar (1959), which decomposes the joint conditional distribution function of $\boldsymbol{Y} = [y_1, \cdots, y_n]'$ conditional on X as

$$\boldsymbol{Y} \mid X \sim H(. \mid X) = C\left(F_1(. \mid X), \cdots, F_n(. \mid X)\right),$$

where $F_t(. \mid X)$ for $t = 1, \cdots, n$ are uniformly distributed marginals - i.e. $F_t(. \mid X) := u_t \sim U[0, 1]$. Note that the elements of $\boldsymbol{Y}$ are uncorrelated, yet exhibit serial dependence which is captured by the copula $C(.)$. The issue with specifying a copula for $\boldsymbol{Y}$ is that the no serial correlation assumption implies an identity correlation matrix. As a result, in the literature, the means of allowing for non-linear serial dependence for processes which are linearly unrelated is often accompanied by assuming that $\boldsymbol{Y}$ conditional on X is distributed as a multivariate Student's t distribution - i.e. $\boldsymbol{Y} \mid X \sim t_\nu(0, I)$, where I is an identity matrix. When I is imposed on the multivariate Student's t distribution, the conditional joint distribution of $\boldsymbol{Y}$ does not factorize into the product of its marginals. Alternatively, we may consider the "jointly symmetric"copulae proposed by Oh and Patton (2016), where the latter can be constructed with any given (possibly asymmetric) copula family. In addition, when they are combined with symmetric marginals, they ensure an identity correlation matrix. A "jointly symmetric"copula is defined as follows

Definition 1 *The n dimensional copula C^{JS}, is jointly symmetric:*

$$C^{JS}(u_1, \cdots, u_n) = \frac{1}{2^n} \sum_{k_1=0}^{2} \cdots \sum_{k_n=0}^{2} (-1)^R C(\tilde{u}_1, \cdots, \tilde{u}_i, \cdots, \tilde{u}_n)$$

$$\textit{where} \quad R = \sum_{i=1}^{n} \mathbf{1}\{k_i = 2\}, \quad \textit{and} \quad \tilde{u}_i = \begin{cases} 1, & k_i = 0 \\ u_i, & k_i = 1 \\ 1 - u_i, & k_i = 2 \end{cases}$$

The general idea is that the average of mirror image rotations of a possibly asymmetric copula along each axis generates a jointly symmetric copula [see Oh and Patton (2016). For instance, the marginals can be assumed to possess standard normal distributions, while the nonlinear dependency is modeled using the jointly symmetric copulae. The Markovian assumption can then be tested by considering in turn, the independent, bivariate, trivariate and higher order multivariate copulae (or multivariate Student's t distributions of varying orders), where the model with the lowest Akaike Information Criterion (AIC hereafter) is then chosen. However, in an extensive simulation analysis, we observe that the order of the Markovianity does not have a significant impact on the power of the test, and as such the assumption that under the alternative hypothesis $\{y_t : t = 1, \cdots, n\}$ and in turn $\{s(y_t) : t = 1, \cdots, n\}$ follow a Markov process of order one is sufficient for testing the null hypothesis of orthogonality.

Now, under assumption **A1**, the probability terms $P\left[y_t \geq 0 \mid \underline{S}_{t-1}, X\right]$ and $P\left[y_t < 0 \mid \underline{S}_{t-1}, X\right]$ in the weight function $a_t(\beta_1)$ can be written as follows:

$$\begin{cases} P\left[y_t \geq 0 \mid \underline{S}_{t-1}, X\right] = P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]^{s(y_{t-1})} P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]^{1-s(y_{t-1})}, \\ \\ P\left[y_t < 0 \mid \underline{S}_{t-1}, X\right] = P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right]^{s(y_{t-1})} P\left[y_t < 0 \mid y_{t-1} < 0, X\right]^{1-s(y_{t-1})}. \end{cases}$$

The use of the above expressions of $P\left[y_t \geq 0 \mid \underline{S}_{t-1}, X\right]$ and $P\left[y_t < 0 \mid \underline{S}_{t-1}, X\right]$ simplifies a lot the calculation of the test statistic $SL_n(\beta_1)$. We have the following result.

Corollary 1 *Under assumptions (1.2) and (1.1), let H_0 and H_1 be defined by (1.3) - (1.4),*

$$\tilde{S}L_n(\beta_1) = \sum_{t=1}^{n} \tilde{a}_t(\beta_1) s(y_t) + \sum_{t=1}^{n} \tilde{b}_t(\beta_1) s(y_t) s(y_{t-1}),$$

where

$$\tilde{a}_1(\beta_1) = \ln\left\{\frac{1 - P\left[\varepsilon_1 < -\beta_1' x_0 \mid X\right]}{P\left[\varepsilon_1 < -\beta_1' x_0 \mid X\right]}\right\}, \quad \tilde{b}_1(\beta_1) = 0$$

and for $t = 2, ..., n$,

$$\tilde{a}_t(\beta_1) = \ln\left\{\frac{1 - \frac{P\left[\varepsilon_t < -\beta_1' x_{t-1},\, \varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}{P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}}{\frac{P\left[\varepsilon_t < -\beta_1' x_{t-1},\, \varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}{P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}}\right\}$$

$$\tilde{b}_t(\beta_1) = \ln\left\{\frac{1 - \left(\frac{P\left[\varepsilon_t < -\beta_1' x_{t-1} \mid X\right]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]} - \frac{P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2},\, \varepsilon_t < -\beta_1' x_{t-1} \mid X\right]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}\right)}{\frac{P\left[\varepsilon_t < -\beta_1' x_{t-1} \mid X\right]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]} - \frac{P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2},\, \varepsilon_t < -\beta_1' x_{t-1} \mid X\right]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}}\right\} - \ln\left\{\frac{1 - \frac{P\left[\varepsilon_t < -\beta_1' x_{t-1},\, \varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}{P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}}{\frac{P\left[\varepsilon_t < -\beta_1' x_{t-1},\, \varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}{P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}}\right\}$$

and suppose the constant $\tilde{c}_1(\beta_1)$ satisfies $P\left[\sum_{t=1}^{n} \tilde{a}_t(\beta_1)\ s(y_t) + \sum_{t=1}^{n} \tilde{b}_t(\beta_1)\ s(y_t) s(y_{t-1}) > \tilde{c}_1(\beta_1)\right] = \alpha$ under H_0, with $0 < \alpha < 1$. Then the test that rejects H_0 when

$$\tilde{S}L_n(\beta_1) > \tilde{c}_1(\beta_1) \tag{1.7}$$

is most powerful for testing H_0 against H_1 among level-α tests based on the signs $\big(s(y_1), ..., s(y_n)\big)'$.

Proof: *See Appendix.*

Now, the calculation of the test statistic $\tilde{S}L_n(\beta_1)$ depends on the univariate and bivariate distributions $P\,[\varepsilon_t < \cdot \mid X]$ and $P\,[\ \varepsilon_{t-1} < \cdot,\ \varepsilon_t < \cdot \mid X]$,

Observe that under the null hypothesis, the signs $s(y_1), ..., s(y_n)$ are i.i.d. according to a Bernoulli $Bi(1, 0.5)$. Thus, the distribution of the test statistic $\tilde{S}L_n(\beta_1)$ only depends on the known weights $\tilde{a}_t(\beta_1)$ and $\tilde{b}_t(\beta_1)$ and does not involve any nuisance parameter under the null hypothesis. Nonparametric assumption (1.2) implies that tests based on $\tilde{S}L_n(\beta_1)$, such as the test given by (1.7), are distribution-free and robust against heteroskedasticity of unknown form, and thus, a nonparametric *pivotal function*. Under the alternative hypothesis, however, the power function of the test depends on the form of the distributions $P\,[\varepsilon_t < . \mid X]$ and $P\,[\ \varepsilon_{t-1} < \cdot,\ \varepsilon_t < \cdot \mid X]$.

A special case is where $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{n-1}, \varepsilon_n$ are distributed according to $N(0, 1)$. As suggested before, since the form of the serial dependence of the errors is non-linear, we may calculate the bivariate probabilities using "jointly-symmetric"copulae by considering the Archimedean Frank, Clayton or Gumbel as the copula family [see Joe (2014)]. Alternatively, we may evaluate the bivariate probabilities $P\,[\ \varepsilon_{t-1} < \cdot,\ \varepsilon_t < \cdot \mid X]$ using a multivariate Student's t distribution by imposing the identity matrix I. Then the optimal test statistic takes the form

$$\tilde{S}L_n(\beta_1) = \sum_{t=1}^{n} \tilde{a}_t(\beta_1) s(y_t) + \sum_{t=1}^{n} \tilde{b}_t(\beta_1) s(y_t) s(y_{t-1}),$$

where

$$\tilde{a}_1(\beta_1) = \ln\left\{\frac{\Phi(\beta_1' x_0)}{1 - \Phi(\beta_1' x_0)}\right\},\ \tilde{b}_1(\beta_1) = 0$$

and for $t = 2, ..., n$,

$$\tilde{a}_t(\beta_1) = \ln\left\{\frac{1 - \frac{C^{JS}(\Phi(-\beta_1' x_{t-1}), \Phi(-\beta_1' x_{t-2}))}{1-\Phi\left(\beta_1' x_{t-2}\right)}}{\frac{C^{JS}(\Phi(-\beta_1' x_{t-1}), \Phi(-\beta_1' x_{t-2}))}{1-\Phi\left(\beta_1' x_{t-2}\right)}}\right\}$$

$$\tilde{b}_t(\beta_1) = \ln\left\{\frac{1 - \left(\frac{1-\Phi(\beta_1' x_{t-1})}{\Phi\left(\beta_1' x_{t-2}\right)} - \frac{C^{JS}(\Phi(-\beta_1' x_{t-1}), \Phi(-\beta_1' x_{t-2}))}{\Phi\left(\beta_1' x_{t-2}\right)}\right)}{\frac{1-\Phi\left(\beta_1' x_{t-1}\right)}{\Phi\left(\beta_1' x_{t-2}\right)} - \frac{C^{JS}(\Phi(-\beta_1' x_{t-1}), \Phi(-\beta_1' x_{t-2}))}{\Phi\left(\beta_1' x_{t-2}\right)}}\right\} - \ln\left\{\frac{1 - \frac{C^{JS}(\Phi(-\beta_1' x_{t-1}), \Phi(-\beta_1' x_{t-2}))}{1-\Phi\left(\beta_1' x_{t-2}\right)}}{\frac{C^{JS}(\Phi(-\beta_1' x_{t-1}), \Phi(-\beta_1' x_{t-2}))}{1-\Phi\left(\beta_1' x_{t-2}\right)}}\right\}$$

where $\Phi(.)$ is the standard normal distribution function and $C^{JS}(u_1, u_2)$ is the bivariate "jointly symmetric"copula with uniformly distributed margins.

The distribution of $\tilde{S}L_n(\beta_1)$, can be simulated under the null hypothesis and the relevant critical values can be evaluated to any degree of precision with a sufficient number of replications. Since the test statistic $\tilde{S}L_n(\beta_1)$ is a continuous random variable, its quantiles are easy to compute. To simulate the distribution of $\tilde{S}L_n(\beta_1)$, the following algorithm is implemented:

1. Compute the test statistic $\tilde{S}L_n(\beta_1)$ based on the observed data, say $\tilde{S}L_n^0(\beta_1)$;
2. Generate a sample $\{y_t\}_{t=1}^n$ of length n under the null H_0 and compute $\tilde{S}L_n^j(\beta_1)$ using that generated sample;
3. Choose B such that $\alpha(B+1)$ is an integer and repeat steps 1-2 B times;
4. Computer the $(1-\alpha)\%$ quantile, say $\tilde{c}_1(\beta_1)$, of the sequence $\{\tilde{S}L_n^j(\beta_1)\}_{j=1}^B$;
5. Reject the null hypothesis at level α if $\tilde{S}L_n^0(\beta_1) \geq c(\beta_1)$.

1.2.2 Testing general full coefficient hypotheses in non-linear regressions

We now consider a non-linear regression model

$$y_t = f(x_{t-1}, \beta) + \varepsilon_t, \;\; t = 1, ..., n, \tag{1.8}$$

where x_{t-1} is an observable $(k+1) \times 1$ vector of stochastic explanatory variables, such that $x_{t-1} = [1, x_{1,t-1}, ..., x_{k,t-1}]'$, $f(\cdot)$ is a scalar function, $\beta \in \mathbb{R}^{(k+1)}$ is an unknown vector of parameters and

$$\varepsilon_t \mid X \sim F_t(. \mid X)$$

where as before $F_t(. \mid X)$ is a distribution function and $X = [x_0, ..., x_{n-1}]$ is an $n \times (k+1)$ matrix. Once again, we suppose that the error terms process $\{\varepsilon_t, t = 1, 2, \cdots\}$ is a strict conditional mediangale, such that

$$P[\varepsilon_t > 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = P[\varepsilon_t < 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = \frac{1}{2}, \tag{1.9}$$

with

$$\boldsymbol{\varepsilon}_0 = \{\emptyset\}, \quad \boldsymbol{\varepsilon}_{t-1} = \{\varepsilon_1, \cdots, \varepsilon_{t-1}\}, \quad \text{for} \quad t \geq 2$$

and where (2.20) entails that $\varepsilon_t \mid X$ has no mass at zero, *i.e.* $P[\varepsilon_t = 0 \mid X]$=0 for all t. We do not require that the parameter vector β be identified. We consider the problem of testing the null hypothesis

$$H(\beta_0) : \beta = \beta_0 \tag{1.10}$$

against the alternative hypothesis

$$H(\beta_1) : \beta = \beta_1. \tag{1.11}$$

A test for $H(\beta_0)$ against $H(\beta_1)$ can be constructed as in Section 1.2.1. First, we note that model (1.8) is equivalent to the transformed model

$$\tilde{y}_t = g(x_{t-1},\, \beta,\, \beta_0) + \varepsilon_t,$$

where $\tilde{y}_t = y_t - f(x_{t-1}, \beta_0)$ and $g(x_{t-1}, \beta, \beta_0) = f(x_{t-1}, \beta) - f(x_{t-1}, \beta_0)$. Thus, testing $H(\beta_0)$ against $H(\beta_1)$ is equivalent to testing

$$\bar{H}_0 : g(x_{t-1},\, \beta, \beta_0) = 0, \text{ for } t = 1, ..., n,$$

against

$$\bar{H}_1 : g(x_{t-1}, \beta, \beta_0) = f(x_t, \beta_1) - f(x_t, \beta_0), \text{ for } t = 1, ..., n.$$

For $\tilde{U}(n) = (s(\tilde{y}_1), ..., s(\tilde{y}_n))'$, where, for $1 \leq t \leq n$,

$$s(\tilde{y}_t) = \begin{cases} 1, \, if \;\, \tilde{y}_t \geq 0 \\ 0, \, if \;\, \tilde{y}_t < 0 \end{cases},$$

the likelihood function of new random sample $\{s(\tilde{y})_t\}_{t=1}^{n}$ conditional on X is given by:

$$L\left(\tilde{U}(n), \beta\right) = P\left[s(\tilde{y}_1) = \tilde{s}_1, ..., s(\tilde{y}_n) = \tilde{s}_n \mid X\right] = \prod_{t=1}^{n} P\left[s(\tilde{y}_t) = \tilde{s}_t \mid \tilde{\mathrm{S}}_{t-1}, X\right],$$

with

$$\tilde{\mathrm{S}}_0 = \{\emptyset\},\ \tilde{\mathrm{S}}_{t-1} = \{s(\tilde{y}_1) = \tilde{s}_1, ..., s(\tilde{y}_{t-1}) = \tilde{s}_{t-1}\},\ \text{for } t \geq 2,$$

and

$$P\left[s(\tilde{y}_1) = \tilde{s}_1 \mid \tilde{\mathrm{S}}_0, X\right] = P\left[s(\tilde{y}_1) = \tilde{s}_1 \mid X\right],$$

where each $\tilde{s}_t$, for $1 \leq t \leq n$, takes two possible values 0 and 1. Thus, we can use the result of proposition 1 to derive a sign-based test for the null hypothesis $H(\beta_0)$ against the alternative hypothesis $H(\beta_1)$, which leads to the following proposition:

Proposition 2 *Under assumptions (1.2) and (1.8),let $H(\beta_0)$ and $H(\beta_1)$ be defined by* (1.10) - (1.11),

$$SN_n(\beta_0|\beta_1) = \sum_{t=1}^{n} a_t(\beta_0|\beta_1) s\left(y_t - f(x_{t-1}, \beta_0)\right)$$

where, for $t = 1, ..., n$,

$$a_t(\beta_0|\beta_1) = \ln\left\{\frac{P\left[\tilde{y}_t \geq 0 \mid \tilde{S}_{t-1}, X\right]}{P\left[\tilde{y}_t < 0 \mid \tilde{S}_{t-1}, X\right]}\right\},$$

and suppose the constant $c_1(\beta_0,\ \beta_1)$ satisfies $P\left[\sum_{t=1}^{n} a_t(\beta_0|\beta_1)\ s\left(y_t - f(x_t, \beta_0)\right) > c_1(\beta_0,\ \beta_1)\right] = \alpha$ under $H(\beta_0)$, with $0 < \alpha < 1$. Then the test that rejects $H(\beta_0)$ when

$$SN_n(\beta_0|\beta_1) > c_1(\beta_0,\ \beta_1)$$

is most powerful for testing $H(\beta_0)$ against $H(\beta_1)$ among level-α tests based on the signs $\left(s(\tilde{y}_1), ..., s(\tilde{y}_n)\right)'$.

As in Section 1.2.1, to make the calculation of the weight function $a_t(\beta_0|\beta_1)$ that depends on the terms $P\left[\tilde{y}_t \geq 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]$ and $P\left[\tilde{y}_t < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]$ feasible, we consider the following assumption.

Assumption A2: *Let $\{\tilde{y}_t, t = 0, 1, \cdots\}$ follow a Markov process of order one. Then under the* *alternative hypothesis, the sign process* $\{s(\tilde{y}_t)\}_{t=0}^{\infty}$ *is a Markov process of the same order.*

Under Assumption **A2**, the terms $P\left[\tilde{y}_t \geq 0 \mid \tilde{\mathrm{S}}_{t-1}, X\right]$ and $P\left[\tilde{y}_t < 0 \mid \tilde{\mathrm{S}}_{t-1}, X\right]$ simplify and can be expressed as follows:

$$\begin{cases} P\left[\tilde{y}_t \geq 0 \mid \tilde{\mathrm{S}}_{t-1}, X\right] = P\left[\tilde{y}_t \geq 0 \mid \tilde{y}_{t-1} \geq 0, X\right]^{s(\tilde{y}_{t-1})} P\left[\tilde{y}_t \geq 0 \mid \tilde{y}_{t-1} < 0, X\right]^{1-s(\tilde{y}_{t-1})}, \\ \\ P\left[\tilde{y}_t < 0 \mid \tilde{\mathrm{S}}_{t-1}, X\right] = P\left[\tilde{y}_t < 0 \mid \tilde{y}_{t-1} \geq 0, X\right]^{s(\tilde{y}_{t-1})} P\left[\tilde{y}_t < 0 \mid \tilde{y}_{t-1} < 0, X\right]^{1-s(\tilde{y}_{t-1})}. \end{cases}$$

The use of the new expressions of the probabilities $P\left[\tilde{y}_t \geq 0 \mid \tilde{\mathrm{S}}_{t-1}, X\right]$ and $P\left[\tilde{y}_t < 0 \mid \tilde{\mathrm{S}}_{t-1}, X\right]$ simplifies the calculation of the test statistic $SN_n(\beta_0|\beta_1)$. We have the following results:

Corollary 2 *Under assumptions (1.2) and (1.1), let* $H(\beta_0)$ *and* $H(\beta_1)$ *be defined by* (1.10) - (1.11),

$$\widehat{SN}_n(\beta_0|\beta_1) = \sum_{t=1}^{n} \tilde{a}_t(\beta_0|\beta_1) s(y_t - f(x_{t-1}, \beta_0)) + \sum_{t=1}^{n} \tilde{b}_t(\beta_1)\ s(y_t - f(x_{t-1}, \beta_0)) s(y_{t-1} - f(x_{t-2}, \beta_0)),$$

where

$$\tilde{a}_1(\beta_0|\beta_1) = \ln\left\{\frac{1 - P\left[\varepsilon_1 < f(x_0, \beta_0) - f(x_0, \beta_1) \mid X\right]}{P\left[\varepsilon_1 < f(x_0, \beta_0) - f(x_0, \beta_1) \mid X\right]}\right\},\ \tilde{b}_1(\beta_0|\beta_1) = 0$$

and for $t = 2, ..., n$,

$$\tilde{a}_t(\beta_0|\beta_1) = \ln\left\{\frac{1 - \frac{P[\varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1),\ \varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X]}{P[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X]}}{\frac{P[\varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1),\ \varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X]}{P[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X]}}\right\}$$

$$\tilde{b}_t(\beta_0|\beta_1) = \ln\left\{\frac{1 - \left(\frac{P\left[\varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1) | X\right]}{1 - P\left[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]} - \frac{P\left[\ \varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1),\ \varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1) | X\right]}{1 - P\left[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]}\right)}{\frac{P\left[\varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1) | X\right]}{1 - P\left[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]} - \frac{P\left[\ \varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1),\ \varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1) | X\right]}{1 - P\left[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]}}\right\}$$

$$- \ln\left\{\frac{1 - \frac{P\left[\varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1),\ \varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]}{P\left[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]}}{\frac{P\left[\varepsilon_t < f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1),\ \varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]}{P\left[\varepsilon_{t-1} < f(x_{t-2},\beta_0) - f(x_{t-2},\beta_1) | X\right]}}\right\}$$

and suppose the constant $\tilde{c}_1(\beta_0, \beta_1)$ *satisfies* $P\left[\widehat{SN}_n(\beta_0|\beta_1) > \tilde{c}_1(\beta_0, \beta_1)\right] = \alpha$ *under* $H(\beta_0)$*, with*

$0 < \alpha < 1$. *Then the test that rejects* $H(\beta_0)$ *when*

$$\widehat{SN}_n(\beta_0|\beta_1) > \tilde{c}_1(\beta_0, \beta_1)$$

is most powerful for testing $H(\beta_0)$ *against* $H(\beta_1)$ *among level-*α *tests based on the signs* $\big(s(\tilde{y}_1), ..., s(\tilde{y}_n)\big)'$.

If we consider a linear function $f(x'_{t-1}, \beta) = \beta' x_{t-1}$, as before we may suppose that ε_t for $t = 1, \cdots, n$ follow $N(0,1)$, which allows us to evaluate the bivariate probabilities by utilizing the "jointly symmetric"copula or the multivariate Student's t distribution with the identity matrix imposed. Then the test statistic for the null hypothesis $H(\beta_0)$ against the alternative $H(\beta_1)$ is given by

$$\widehat{SN}_n(\beta_0|\beta_1) = \sum_{t=1}^{n} \tilde{a}_t(\beta_0|\beta_1)s(y_t - \beta_0' x_{t-1}) + \sum_{t=1}^{n} \tilde{b}_t(\beta_1)s(y_t - \beta_0' x_{t-1})s(y_{t-1} - \beta_0' x_{t-2}),$$

where

$$\tilde{a}_1(\beta_0|\beta_1) = \ln\left\{\frac{\Phi\left((\beta_1-\beta_0)' x_0\right)}{1-\Phi\left((\beta_1-\beta_0)' x_0\right)}\right\}, \quad \tilde{b}_1(\beta_0|\beta_1) = 0,$$

and for $t = 2, ..., n$,

$$\tilde{a}_t(\beta_0|\beta_1) = \ln\left\{\frac{1-\frac{C^{JS}(\Phi((\beta_0-\beta_1)' x_{t-1}), \Phi((\beta_0-\beta_1)' x_{t-2}))}{1-\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)}}{\frac{C^{JS}(\Phi((\beta_0-\beta_1)' x_{t-1}), \Phi((\beta_0-\beta_1)' x_{t-2}))}{1-\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)}}\right\},$$

$$\tilde{b}_t(\beta_0|\beta_1) = \ln\left\{\frac{1-\left(\frac{1-\Phi\left((\beta_1-\beta_0)' x_{t-1}\right)}{\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)} - \frac{C^{JS}(\Phi((\beta_0-\beta_1)' x_{t-1}), \Phi((\beta_0-\beta_1)' x_{t-2}))}{\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)}\right)}{\frac{1-\Phi\left((\beta_1-\beta_0)' x_{t-1}\right)}{\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)} - \frac{C^{JS}(\Phi((\beta_0-\beta_1)' x_{t-1}), \Phi((\beta_0-\beta_1)' x_{t-2}))}{\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)}}\right\}$$

$$-\ln\left\{\frac{1-\frac{C^{JS}(\Phi((\beta_0-\beta_1)' x_{t-1}), \Phi((\beta_0-\beta_1)' x_{t-2}))}{1-\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)}}{\frac{C^{JS}(\Phi((\beta_0-\beta_1)' x_{t-1}), \Phi((\beta_0-\beta_1)' x_{t-2}))}{1-\Phi\left((\beta_1-\beta_0)' x_{t-2}\right)}},\right\}$$

where $\Phi(.)$ is the standard normal distribution function and $C^{JS}(u_1, u_2)$ is the "jointly-symmetric"bivariate copula with uniformly distributed marginals. As in Section 1.2.1, the test statistic $\widehat{SN}_n(\beta_0|\beta_1)$ depends on a particular alternative hypothesis β_1. In practice, the latter is unknown, which makes the proposed POS test infeasible. However, in Section 1.3 we will suggest an adaptive approach [see Dufour and Taamouti (2010a)] which can be used to choose an optimal alternative hypothesis at which the power of the test is maximized.

1.3 Choice of the optimal alternative hypothesis

Point-optimal tests depend on the alternative $\beta = \beta_1$, which in practice is unknown. Formally, the test statistic $\widehat{SN}_n(\beta_0|\beta_1)$ for testing the linear full-coefficient hypothesis (1.10) is a function of β_1

$$\widehat{SN}_n(\beta_0|\beta_1) = \sum_{t=1}^{n} \tilde{a}_t(\beta_0|\beta_1)s(y_t - x'_{t-1}\beta_0) + \sum_{t=1}^{n} \tilde{b}_t(\beta_1)s(y_t - x'_{t-1}\beta_0)s(y_{t-1} - x'_{t-2}\beta_0),$$

which in turn implies that its power function, say $\Pi(\beta_0, \beta_1)$, is also a function of β_1. Therefore, the choice of the alternative β_1 has a direct impact on its power function. In other words,

$$\Pi(\beta_0, \beta_1) = P[\widehat{SN}_n(\beta_0|\beta_1) > \tilde{c}_1(\beta_0, \beta_1) \mid H(\beta_1)],$$

where $\tilde{c}_1(\beta_0, \beta_1)$ satisfies the constraint

$$P[\widehat{SN}_n(\beta_0|\beta_1) > \tilde{c}_1(\beta_0, \beta_1)|H(\beta_0)] \leq \alpha.$$

Our objective is to choose the value of β_1 at which the power of the POS-based test statistic is maximized and is close to that of the power envelope. This can be accomplished in a number of ways. Dufour and Taamouti (2010a) suggest an adaptive approach based on the split-sample technique [see Dufour and Jasiak (2001)] for estimating the optimal alternative to make size control easier and maximize the power. For a review of adaptive approach for parametric tests with non-standard distributions see Dufour and Taamouti (2003) and Dufour et al. (2008).

This approach consists of splitting the sample into two independent parts, where the alternative β_1 is estimated using the first part, while the POS test-statistic $\widehat{SN}_n(\beta_0|\beta_1)$ is calculated using the second part of the sample, along with the alternative β_1 estimated using the first part. By adopting this technique, size control is easier and the power function of the POS-test traces out the power envelope. Let $n = n_1 + n_2$, $y = (y'_{(1)}, y'_{(2)})'$, $X = (X'_{(1)}, X'_{(2)})'$, and $\varepsilon = (\varepsilon'_{(1)}, \varepsilon'_{(2)})'$, where $y_{(i)}$, $X_{(i)}$ and $u_{(i)}$ for $i \in \{1, 2\}$ each have n_i rows. The first n_1 observations of y and X can thus be denoted by $y_{(1)}$ and $X_{(1)}$, which are used for estimating the alternative hypothesis β_1 with the OLS estimator:

$$\hat{\beta}_{(1)} = (X'_{(1)} X_{(1)})^{-1} X'_{(1)} y_{(1)}.$$

Alternatively, in the case of extreme outliers other robust estimators that are less sensitive to outliers can be utilized [see Maronna et al. (2019) for a review of robust estimators]. Since $\hat{\beta}_{(1)}$ is independent of $X_{(2)}$, the last n_2 observations can be used to calculate the test statistic and obtain a valid POS test

$$\widehat{SN}_n(\beta_0|\beta_{(1)}) = \sum_{t=n_1+1}^{n} \tilde{a}_t(\beta_0|\beta_{(1)}) s(y_t - x'_{t-1}\beta_0) + \sum_{t=n_1+1}^{n} \tilde{b}_t(\beta_{(1)}) s(y_t - x'_{t-1}\beta_0) s(y_{t-1} - x'_{t-2}\beta_0),$$

Different choices for n_1 and n_2 is possible. However, as Dufour and Taamouti (2010a) have noted, the number of observations retained for the first and the second sub-samples has a direct impact on the power of the test, and a more powerful test is obtained when a relatively small number of observations is used for estimating the alternative and more observations are saved for calculating the test statistic. Having conducted a simulation study to compare the power-curves of split-sample-based POS tests to the power envelope, they find that using approximately 10% of the sample to estimate the alternative yields a power which is very close to that of the power envelope. Therefore, we follow Dufour and Taamouti (2010a) by using the first 10% of the sample to estimate the alternative and the remaining 90% to calculate the test statistic.

1.4 POS confidence regions

In this Section, we follow Dufour and Taamouti (2010a) and Coudin and Dufour (2009) to discuss the process of building confidence regions at a given significance level α, say $C_\beta(\alpha)$, for a vector (sub-vector) of the unknown parameters β using the proposed POS test. We consider again the linear regression (1.8) and suppose we wish to test the null hypothesis (1.10) against the alternative hypothesis (1.11). Formally, the idea involves finding all the values of $\beta_0 \in \mathbb{R}^{(k+1)}$ such that

$$\widehat{SN}_n(\beta_0|\,\beta_1) = \sum_{t=1}^{n} \tilde{a}_t(\beta_0 \mid \beta_1) s(y_t - \beta_0' x_{t-1}) + \sum_{t=1}^{n} \tilde{b}_t(\beta_1) s(y_t - \beta_0' x_{t-1}) s(y_{t-1} - \beta_0' x_{t-2}) < \tilde{c}_1(\beta_0, \beta_1) \tag{1.12}$$

where the critical value $\tilde{c}_1(\beta_0, \beta_1)$ satisfies the constraint

$$P\left[\widehat{SN}_n(\beta_0|\,\beta_1) > \tilde{c}_1(\beta_0, \beta_1) | \beta = \beta_0\right] \le \alpha$$

Thus, the confidence region $C_\beta(\alpha)$ of the vector of parameters β is defined as

$$C_\beta(\alpha) = \left\{\beta_0 : \widehat{SN}_n(\beta_0|\,\beta_1) < \tilde{c}_1(\beta_0, \beta_1) | P[\widehat{SN}_n(\beta_0|\,\beta_1) > \tilde{c}_1(\beta_0, \beta_1) | \beta = \beta_0] \le \alpha\right\}.$$

Given the confidence region $C_\beta(\alpha)$, confidence intervals for the components of vector β can be obtained using the projection techniques. Confidence sets in the form of transformations T of $\beta \in \mathbb{R}^m$, $T(C_\beta(\alpha))$ for $m \le (k+1)$ can easily be found using the said techniques. Since, for any set $C_\beta(\alpha)$

$$\beta \in C_\beta(\alpha) \implies T(\beta) \in T(C_\beta(\alpha)), \tag{1.13}$$

we have

$$P[\beta \in C_\beta(\alpha)] \ge 1 - \alpha \implies P[T(\beta) \in T(C_\beta(\alpha))] \ge 1 - \alpha, \tag{1.14}$$

where

$$T(C_\beta(\alpha)) = \{\delta \in \mathbb{R}^m : \exists \beta \in C_\beta(\alpha), T(\beta) = \delta\}.$$

From (1.13) and (1.14), it is evident that the set $T(C_\beta(\alpha))$ is a conservative confidence set for $T(\beta)$ with level $1 - \alpha$. If $T(\beta)$ is a scalar, then we have

$$P[\inf\{T(\beta_0),\ \text{ for }\ \beta_0 \in C_\beta(\alpha)\} \le T(\beta) \le \sup\{T(\beta_0),\ \text{ for }\ \beta_0 \in C_\beta(\alpha)\}] > 1 - \alpha.$$

To obtain valid conservative confidence intervals for the individual component β_j in (1.8) under assumption (1.2), we follow Coudin and Dufour (2009) by implementing a global numerical optimization search algorithm to solve the problem

$$\min_{\beta\in\mathbb{R}^{(k+1)}} \beta_j \;\; s.c. \;\; \widehat{SN}_n(\beta_0 \mid \beta_1) < \tilde{c}_1(\beta_0,\beta_1), \qquad \max_{\beta\in\mathbb{R}^{(k+1)}} \beta_j \;\; s.c. \;\; \widehat{SN}_n(\beta_0 \mid \beta_1) < \tilde{c}_1(\beta_0,\beta_1) \quad (1.15)$$

where the critical value $c(\beta_0, \beta_1)$ at level α, is computed using B replications of the statistic $\widehat{SN}_n^{(i)}(\beta_0 \mid \beta_1)$ under the null hypothesis and finding its $(1-\alpha)$ quantile. Using the projection technique, multiple tests maintain control of the overall level when performed on an arbitrary number of hypotheses.

1.4.1 Numerical illustration

Following Coudin and Dufour (2009), we illustrate the projection technique by generating a process with sample size $n = 500$, such that

$$y_t = \beta_0 + \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1} + \varepsilon_t \overset{i.i.d}{\sim} \begin{cases} N(0,1) & \text{with probability } 0.95 \\ N(0,100^2) & \text{with probability } 0.05 \end{cases}$$

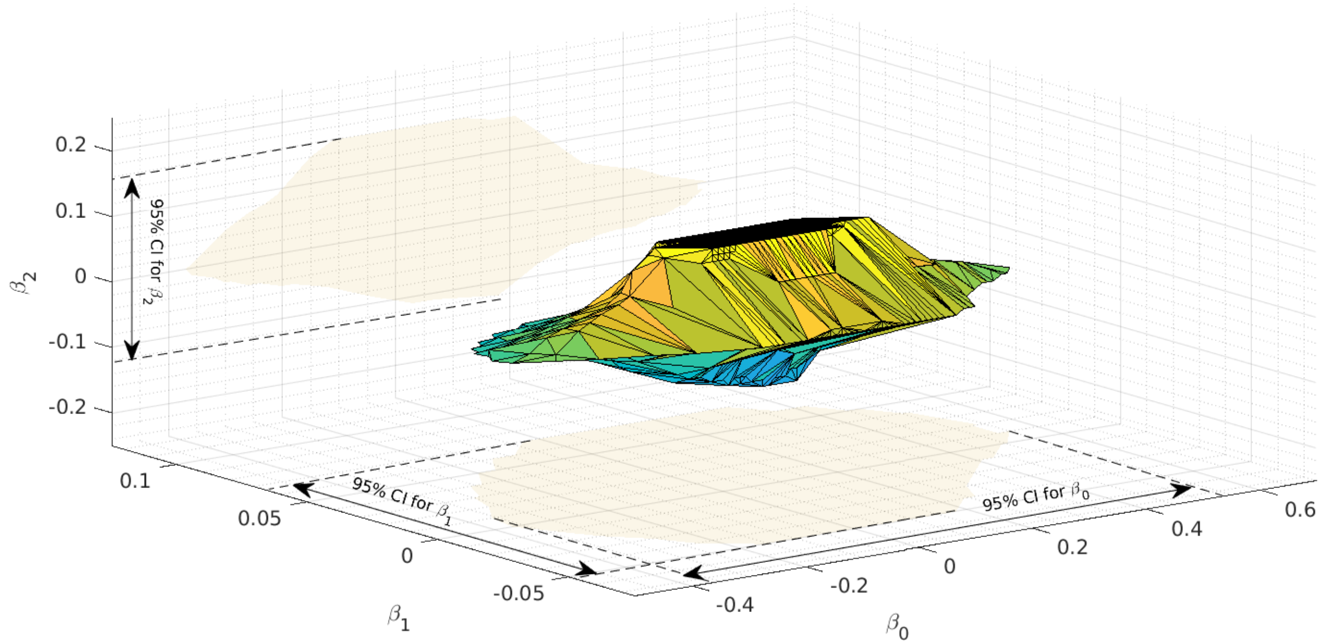
where $\beta_0 = \beta_1 = \beta_2 = 0$ and

$$x_{1,t} = \theta_1 x_{1,t-1} + u_{1,t}$$
$$x_{2,t} = \theta_2 x_{2,t-1} + u_{2,t}$$

with $\theta_1 = \theta_2 = 0.9$ The initial values of x_1 and x_2 are respectively given by: $x_{1,0} = \frac{u_{1,0}}{\sqrt{1-\theta_1^2}}$ and $x_{2,0} = \frac{u_{2,0}}{\sqrt{1-\theta_2^2}}$, and $u_{1,t}$ and $u_{2,t}$ are generated from $N(0,1)$.

The exact inference procedure is conducted with $B = 1,000$ replications of the test statistic under the null hypothesis. As β is a vector in three-dimensional space, the confidence region and the projections can be illustrated graphically. The tests of $H_0(\beta^*) : \beta = \beta^*$ are performed on a grid for $\beta^* = (\beta_0^*, \beta_1^*, \beta_2^*)$. Due to the curse of dimensionality encountered in the process of creating a grid for the parameters, the *simulated annealing optimization algorithm* is initially used to solve problem (1.15) for each parameter β_i, to obtain a realistic size dimension of the grid [see Goffe et al. (1994) for a review of the simulated annealing algorithm].

Figure 1.1: 95% confidence region for the unknown vector $\beta = (\beta_0, \beta_1, \beta_2)$ obtained by searching a three-dimensional grid β^* using the 10% SS-POS test.



Note: The shaded regions on the $\beta_0 - \beta_1$ and $\beta_2 - \beta_1$ planes are the shadows casted by the three-dimensional confidence region, which simplify the visual identification of the 95% confidence intervals for each parameter β_i.

The optimizations were performed using MATLAB software on a high-performance computing (HPC) cluster, by utilizing six nodes each equipped with Intel(R) Xeon(R) 16-core processors (2.40GHz). The simulated annealing algorithm's speed of adjustment was set to 0.25, with a temperature reduction factor of 75%, an initial temperature of 50 and a convergence criteria of 0.01. All algorithms converged in less than an hour. Once the global maximum and minimum for each parameter β_i were obtained, the grid was constructed by the Cartesian product of the linearly spaced distance between the β_i's maxima and minima.

Table 1.1: Comparison of the 95% confidence intervals obtained for the unknown parameters β_0, β_1 and β_2 using the 10% SS-POS-test, with those achieved using the T-test and T-test based on White (1980) variance correction.

		OLS	White	10% SS-POS
β_0	95% CI	**[-0.01, -0.00]**	$[-0.01, 0.00]$	$[-0.37, 0.55]$
β_1	95% CI	**[-1.04, -0.60]**	**[-1.09, -0.56]**	$[-0.05, 0.07]$
β_2	95% CI	**[0.47, 0.67]**	**[0.45, 0.69]**	$[-0.12, 0.16]$

Note: The confidence intervals in bold do not contain the value of zero and imply significance at the 5% level.

It is evident that the 10% split-sample POS-based test outperforms the T-test and the T-test based on white (1980) variance correction test, as the former correctly fails to reject the null hypothesis of orthogonality at the 5% level, whereas the latter two tests reject the null hypothesis in favor of the alternative for almost all parameters.

1.5 Monte Carlo study

In this Section, we provide simulation results that illustrate the performance of the POS-based tests proposed in the previous Sections. We have limited our results to two groups of data generating processes (DGPs) which correspond to different symmetric and asymmetric distributions and different forms of heteroskedasticity.

1.5.1 Simulation setup

We assess the performance of the proposed 10% SS-POS tests in terms of size and power, by considering various DGPs with symmetric and asymmetric distributions and different forms of heteroskedasticity. The DGPs under consideration are supposed to mimic different scenarios that are often encountered in practical settings within the domains of predictive regressions. The performance of the 10% SS-POS tests is compared to that of a few other tests, by considering the following linear predictive regression model

$$y_t = \beta x_{t-1} + \varepsilon_t \tag{1.16}$$

where β is an unknown parameter. Furthermore, we follow Mankiw and Shapiro (1986) by assuming that x_t is a stationary AR(1) process

$$x_t = \theta x_{t-1} + u_t \tag{1.17}$$

such that u_t are mutually independent, and each u_t is independent of x_{t-k} for $k \geq 1$. Moreover, the disturbances (ε_t, u_t) are distributed as bivariate normal, with the contemporaneous covariance matrix

$$\Sigma_{\varepsilon u} = \begin{bmatrix} 1 & \sigma_{\varepsilon u} \\ \sigma_{\varepsilon u} & \sigma_u^2 \end{bmatrix}$$

Therefore, there is feedback from u_t to x_t through ε_t, which implies that $corr(\varepsilon_t, x_{t+k}) \neq 0$ for $k \geq 0$. Thus, as the disturbance vector $[\varepsilon_1, \cdots, \varepsilon_n]'$ is not independent of the regressor vector $[x_0, \cdots, x_{n-1}]'$, the OLS estimator is biased in finite-samples and the T-statistic has a non-standard distribution. Mankiw and Shapiro (1986) perform an extensive simulations exercise by considering different values of the correlation between u_t and ε_t (say ρ) and find that in small samples, as θ and ρ approach unity, the T-test using asymptotic critical values leads to oversized tests; however, this size distortion improves as $n \to \infty$.

To compare the performance of certain parametric and non-parametric tests to that of the POS-based tests, the data is generated from model (1.16), with the stationary process x_t specified as (1.17) and by further setting

$$u_t = \rho\varepsilon_t + w_t\sqrt{1-\rho^2} \tag{1.18}$$

for $\rho = 0$, 0.1, 0.5, 0.9, where ε_t and w_t are assumed to be independent. The initial value of x is given by: $x_0 = \frac{w_0}{\sqrt{1-\theta^2}}$. Further, w_t are generated from $N(0,1)$ and we assign $\theta = 0.9$.

The errors ε_t are i.n.i.d and are categorized by two groups in our simulation study. In the first group, we consider DGPs where the residuals ε_t possess symmetric and asymmetric distributions:

1. normal distribution: $\varepsilon_t \sim N(0,1)$;

2. Cauchy distribution: $\varepsilon_t \sim Cauchy$;

3. Student t distribution with two degrees of freedom: $\varepsilon_t \sim t(2)$;

4. Mixture of normal and Cauchy distributions: $\varepsilon_t \sim s_t \mid \varepsilon_t^C \mid -(1-s_t) \mid \varepsilon_t^N \mid$, where ε_t^C follows Cauchy distribution, ε_t^N follows $N(0,1)$ distribution, and

$$P\left(s_t = 1\right) = P\left(s_t = 0\right) = \frac{1}{2}.$$

The second group of DGPs represents different forms of heteroskedasticity:

5. break in variance:

$$\varepsilon_t \sim \begin{cases} N(0,1) & \text{for } t \neq 25 \\ \sqrt{1000}N(0,1) & \text{for } t = 25 \end{cases};$$

6. exponential variance: $\varepsilon_t \sim N(0,\sigma_\varepsilon^2(t))$ and $\sigma_\varepsilon(t) = \exp(0.5t)$;

7. GARCH(1, 1) plus jump variance:

$$\sigma_\varepsilon^2(t) = 0.00037 + 0.0888\varepsilon_{t-1}^2 + 0.9024\sigma_\varepsilon^2(t-1)\,,$$

$$\varepsilon_t \sim \begin{cases} N(0,\sigma_\varepsilon^2(t)) & \text{for } t \neq 25 \\ 50N(0,\sigma_\varepsilon^2(t)) & \text{for } t = 25 \end{cases};$$

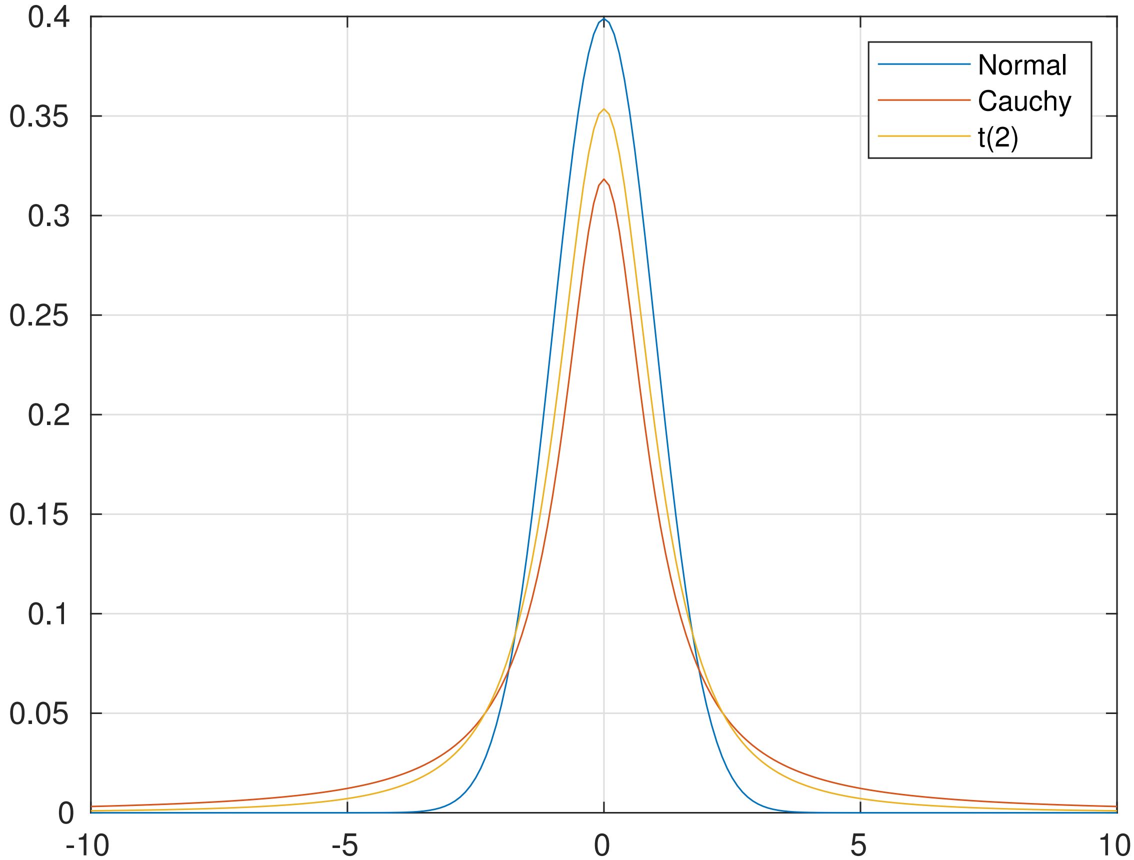
8. nonstationary GARCH(1, 1) variance: $\varepsilon_t \sim N(0,\sigma_\varepsilon^2(t))$ and

$$\sigma_\varepsilon^2(t) = 0.75\varepsilon_{t-1}^2 + 0.75\sigma_\varepsilon^2(t-1)\,.$$

We implement the POS-based test and other tests, which are intended to be robust against heteroskedasticity and non-normality, to test the null hypothesis of orthogonality - i.e. $H_0 : \beta = 0$. As in Dufour and Taamouti (2010a), Monte Carlo simulations are used to compare the size and power of the 10% split-sample POS-based tests (10% SS-POS test hereafter) to those of T-test, T-test based on White (1980) variance correction (hereafter WT-test), and sign-based test proposed by Campbell and Dufour (1995) (CD(1995) test hereafter). The simulation study involves $M_1 = 10,000$ iterations for evaluating the probability distribution of POS test statistic and $M_2 = 5,000$ iterations to estimate the power functions of POS test and other tests. We consider a sample size of $n = 50$ for conducting the simulation exercise. Note that the sign-based test statistic of Campbell and Dufour (1995) possesses a discrete distribution, as a result of which it is not possible (without randomization) to attain test whose size is exactly 5%. In our simulations study, the size of the aforementioned test is 5.95% for $n = 50$.

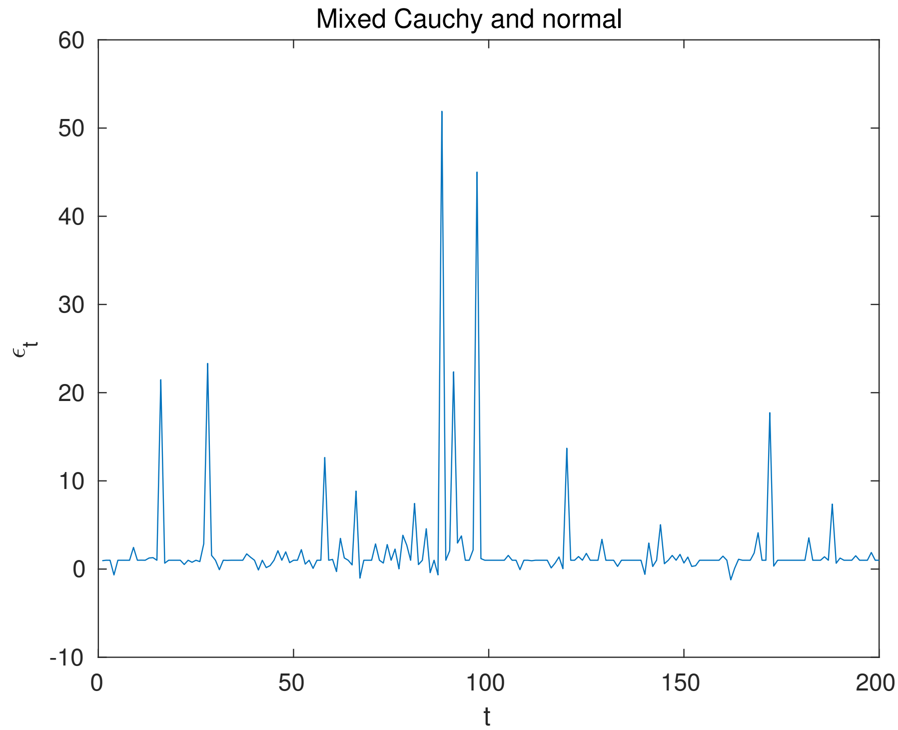
As in Mankiw and Shapiro (1986), it is further possible to consider values of ρ and θ closer to unity at which the size distortions of T-type tests are magnified. For instance, the size of the T-test in their study is shown to be severely distorted with values of $\theta = 0.999$ and $\rho = 1.0$, given a sample size of $n = 50$. The simulations for the latter scenario can be found in the Appendix for standard normal disturbances. It must be noted that as the exact finite-sample distribution of the POS-based tests are simulated, our tests control size regardless of the values of ρ and θ - the results in figure (1.10) confirm these findings. It is further evident that although the size distortions for the T-test and T-test based on White (1980) variance correction improve in large samples, these tests still reject the null hypothesis at twice and thrice their nominal level respectively given a sample of $n = 500$ observations.

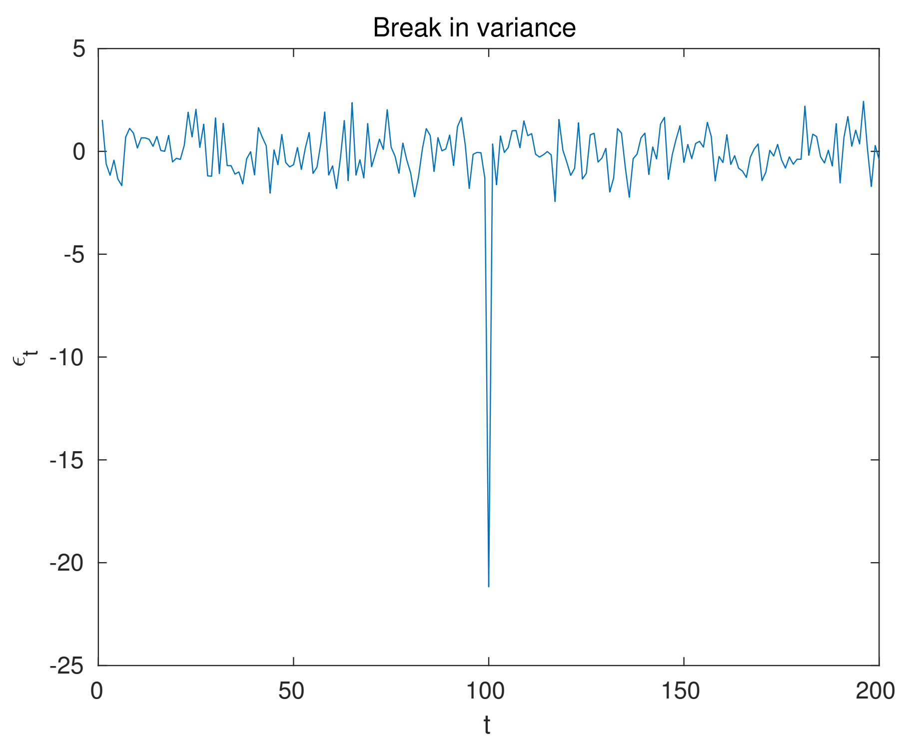
Figure 1.2: Symmetric distributions



Note: In this figure, we compare the symmetric Normal, Cauchy and Student's distribution with two degree of freedom - i.e. $\nu = 2$

Figure 1.3: Time-varying distributions





Note: In this figure, we compare the disturbances generated using the mixed normal and Cauchy distributions, as well as normal distribution with break in variance.

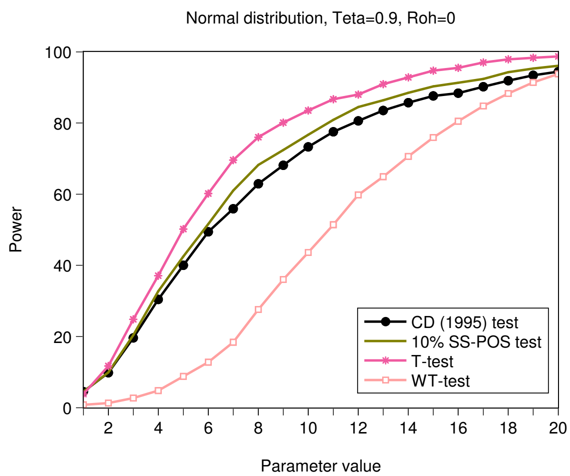
The DGPs considered in this chapter have been inspired by the simulation exercises conducted in the previous studies [see Mankiw and Shapiro (1986), Campbell and Dufour (1995), Coudin and Dufour (2009) and Dufour and Taamouti (2010a)]. The first three DGPs all possess symmetrical distributions that are independent and identical across different observations $t = 1, \cdots, n$. Evidently, as depicted in figure (2.16), the Cauchy and Student's t distribution possess heavier tails in comparison to that of the normal distribution. The standard error of the coefficients are inflated in the presence of heavy tails, as a result of which the power of the T-type tests tend to be poor in comparison to other measures of central tendency (such as the median). Furthermore, the length of the confidence intervals are extended when the data is sampled from heavy tailed distributions. DGP 4 is a mixture of Cauchy and Gaussian distribution; as such, while the errors are independent, they are not identically distributed across different observations [see figure 1.3]. DGP 4 is inspired by Magdalinos and Phillips (2009), who note that when x_t is moderately explosive (with $\theta > 1$), the least squares estimator is mixed normal with Cauchy-type tail behavior with an explosive convergence rate. The second group of DGPs covers different forms of heteroskedasticity, such as conditional heteroskedasticity (e.g. stationary and non-stationary GARCH models) and other forms of non-linear dependencies. Dufour and Taamouti (2010a) show that under certain forms of heteroskedasticity, T-type tests are not valid; hence, these DGPs fit well within the domains of our study.

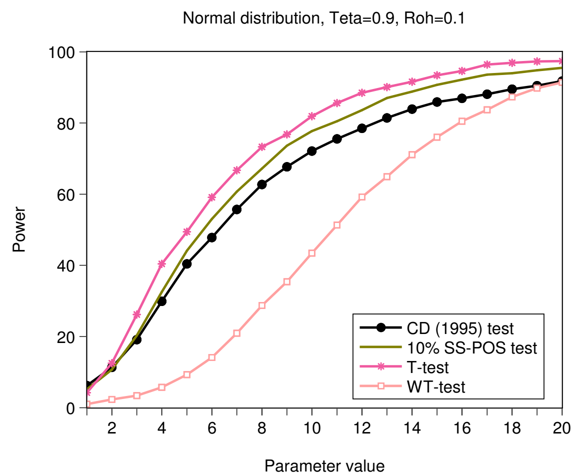
1.5.2 Simulation results

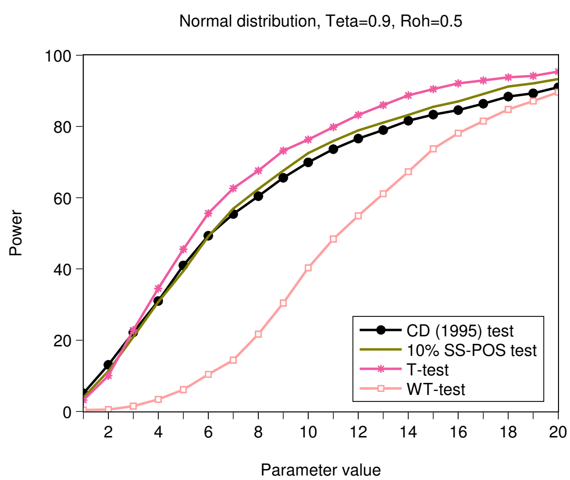
Monte Carlo simulation results are presented in Figures 1.4-1.8. These results correspond to different DGPs described in Section 1.5.1. The figures compare the power of the 10% SS-POS test to the T-test, WT-test, and CD (1995) test. The results are detailed below.

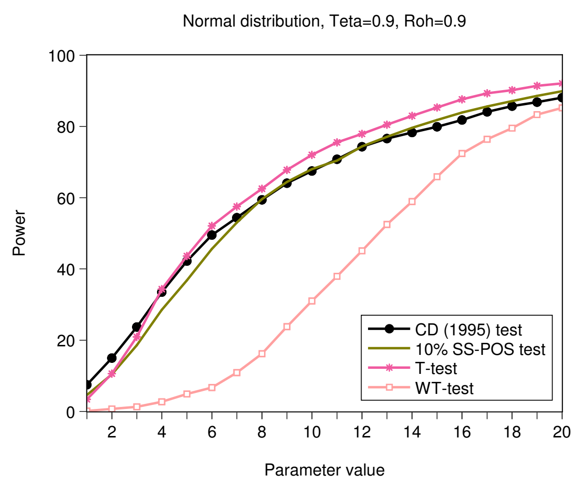
First, Figure 1.4 compares the power function of the above tests in the case where the error term ε_t in the model (1.16) is normally distributed. From this we see that all these tests control size, except WT-test which is undersized. We also find that T-test is more powerful than 10% SS-POS test, CD (1995) test, and WT-test. This result is expected since under normality T-test is the most powerful test. However, the power of 10% SS-POS test has the second best power among the

Figure 1.4: Power comparisons: different tests. Normal error distributions with different values of ρ in (1.18) and $\theta = 0.9$ in (1.17).









Note: These figures compare the power function of the 10% SS-POS test with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

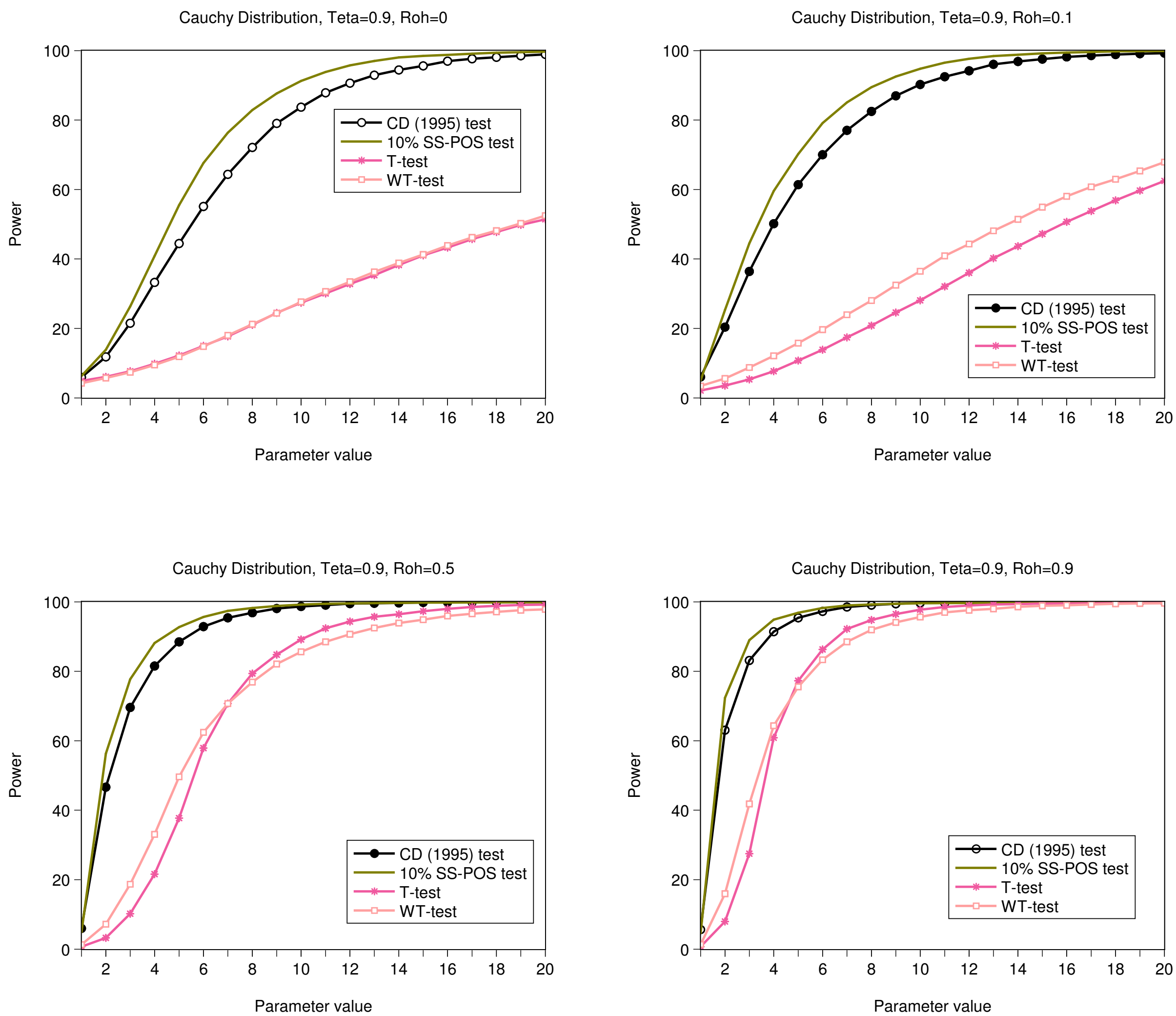
other tests. These results are still the same when we increase the correlation coefficient ρ, except that when there high correlation between the error terms ε_t and w_t the power curves of T-test, 10% SS-POS test and CD (1995) test become closer to each other.

Second, Figure 1.5 corresponds to the cases where the error term ε_t follows Cauchy distribution. From this we see that 10% SS-POS test is more powerful than CD (1995) test, WT-test, and the T-test. It seems that the latter two tests are undersized. 10% SS-POS test and CD (1995) test have much more power than WT-test and T-test for small values (0 and 0.1) of correlation coefficient ρ, but the difference in power decreases when we increase ρ even if it still quite important.

Third, Figure 1.6 corresponds to the cases where the error term ε_t follows a mixture of normal and Cauchy distributions. The results show that 10% SS-POS test is again more powerful than CD (1995), T-test, and the WT-test. The difference in power is much more significant when the correlation coefficient ρ is smaller.

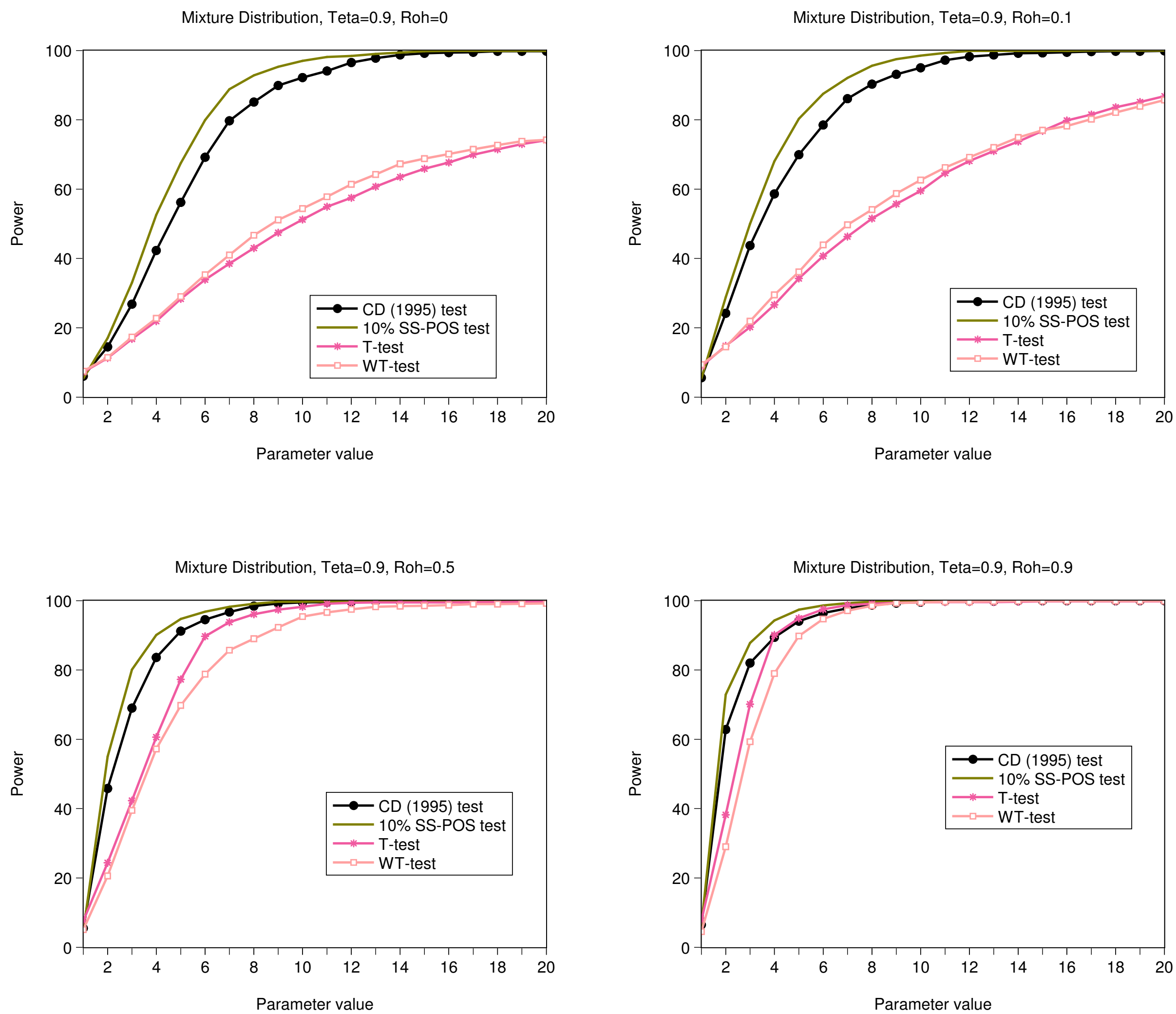
Finally, Figures 1.7 and 1.8 compare the power function of the 10% SS-POS test, CD (1995) test, WT-test, and T-test in the case where ε_t follows normal distribution with a break in variance and an exponential variance, respectively. Figure 1.7 shows that in the presence of break in variance, WT-test and T-test are undesized, whereas 10% SS-POS test and CD (1995) test control size. In addition, 10% SS-POS test has more power that the other tests. The CD (1995) test has the second best power followed by WT-test and T-test. The power of these tests improve when we increase the correlation coefficient ρ. Figure 1.8 shows that in the case of exponential variance, the WT-test, and T-test are oversized. We find that 10% SS-POS test has more power than CD (1995) test when ρ is equal to zero. However, CD (1995) test becomes more powerful than 10% SS-POS test when correlation coefficient ρ increases. The difference in power between the latter two tests becomes small for higher values of ρ.

Figure 1.5: Power comparisons: different tests. Cauchy error distributions with different values of ρ in (1.18) and $\theta = 0.9$ in (1.17).



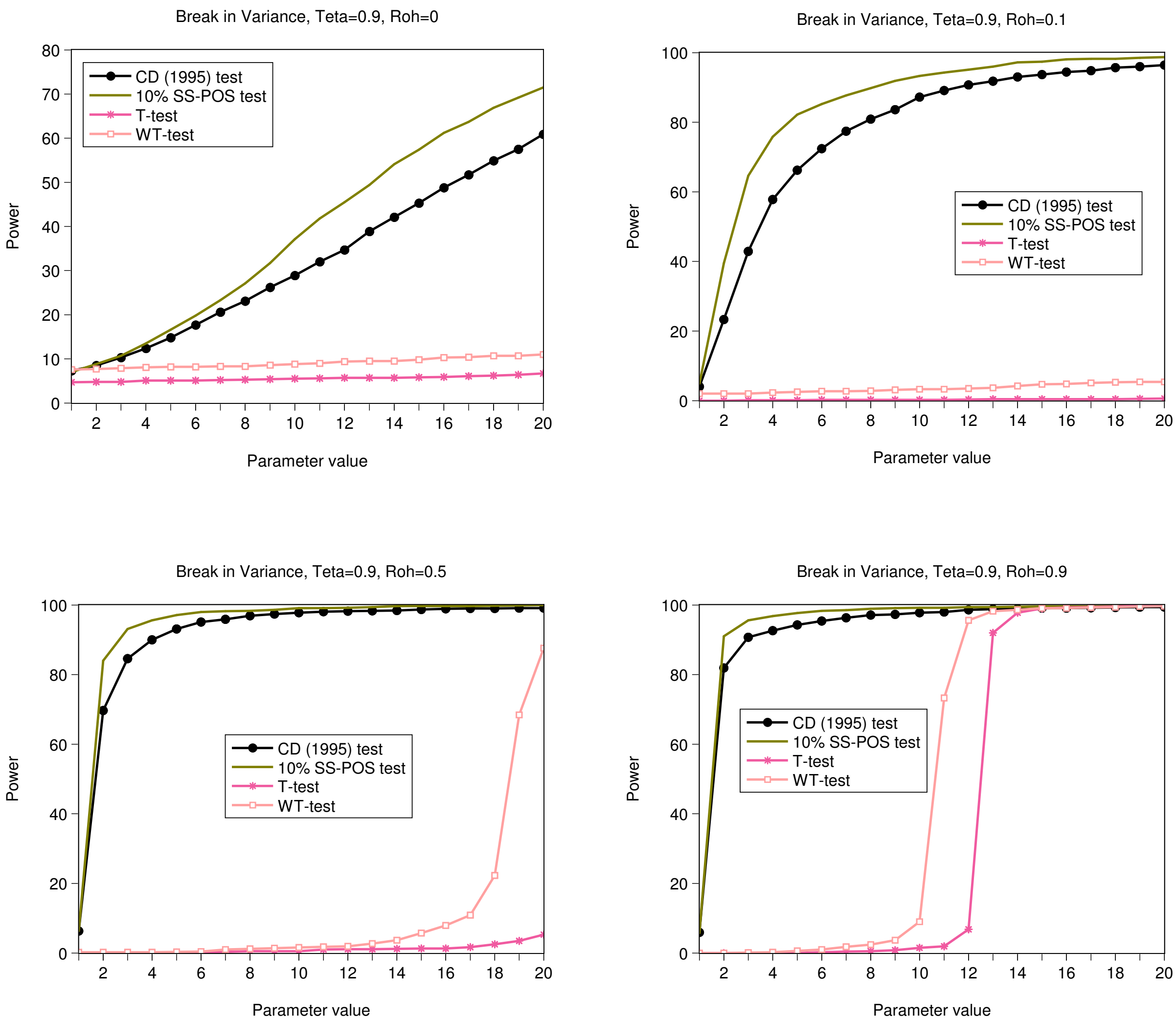
Note: These figures compare the power function of the 10% SS-POS test with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

Figure 1.6: Power comparisons: different tests. Mixture error distributions with different values of ρ in (1.18) and $\theta = 0.9$ in (1.17).



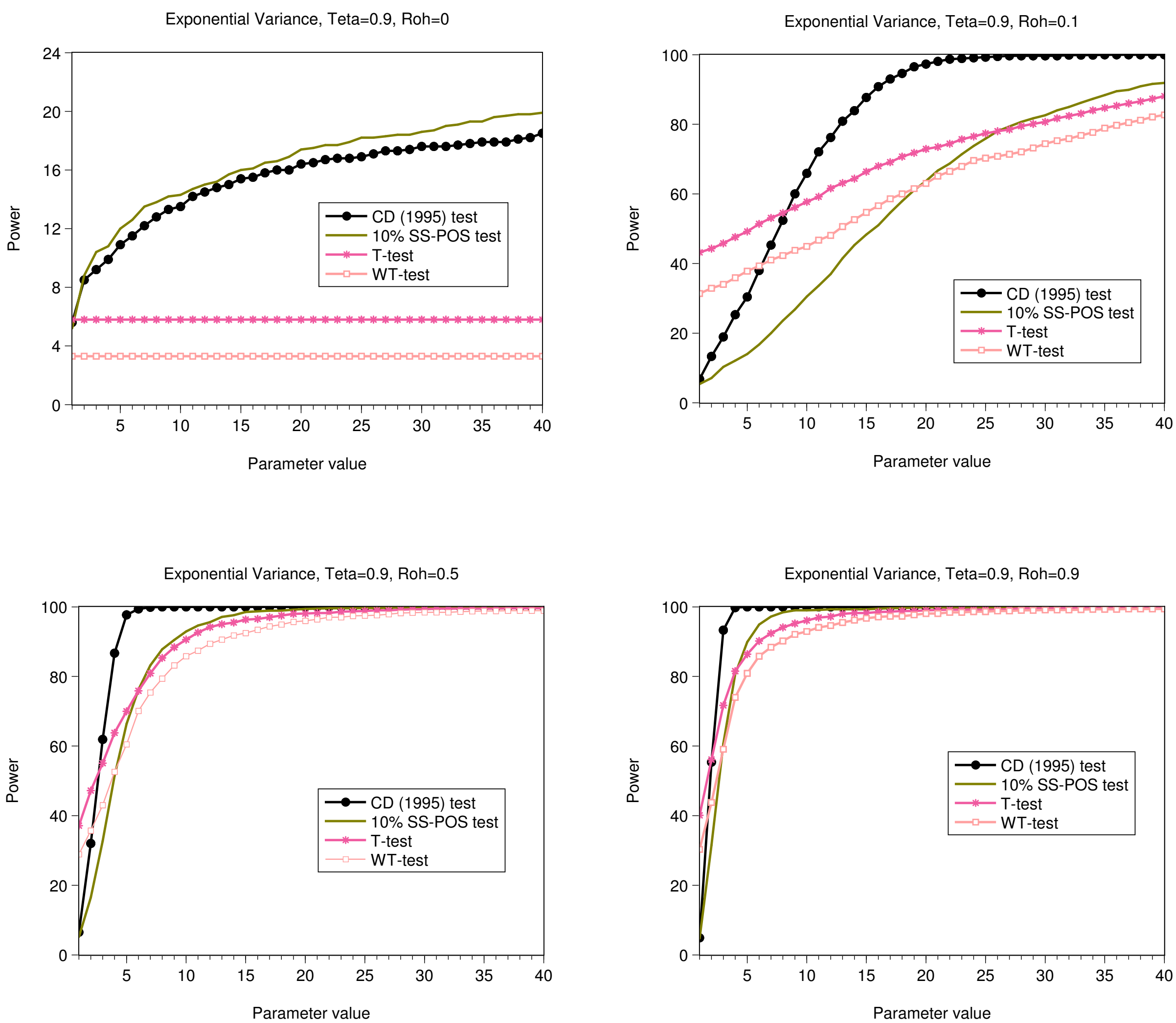
Note: These figures compare the power function of the 10% SS-POS test with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

Figure 1.7: Power comparisons: different tests. Normal error distributions with break in variance, different values of ρ in (1.18) and $\theta = 0.9$ in (1.17).



Note: These figures compare the power function of the 10% SS-POS test with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

Figure 1.8: Power comparisons: different tests. Normal error distributions with Exp(t) variance, different values of ρ in (1.18) and $\theta = 0.9$ in (1.17).



Note: These figures compare the power function of the 10% SS-POS test with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

1.6 Empirical application

In this Section, we consider an empirical application of the proposed 10% SS-POS tests to illustrate its practical relevance. Valuation ratios are widely considered as predictors of stock returns and are generally known to be persistent. Therefore, they fit well within the framework of our study. In what follows, we specifically divert our attention to an application in the context of stock return predictability using the said ratios.

1.6.1 Stock return predictability using valuation ratios

Many studies have investigated the predictive power of valuation ratios on excess stock returns. Dividend-price and earnings-price ratios are among few that were the focus of study in the early 1980s. The attention to these ratios was heightened when Rozeff (1984), Fama and French (1988), and Campbell and Shiller (1988) showed the ratios positive correlation with ex-post stock returns. Fama and French (1988) find that in short horizons dividend yields only explain a small fraction of the variation in time-varying returns, yet in longer horizons (beyond one year) this proportion is significantly increased. Campbell and Shiller (1988) employ a two-variable system approach with the lagged log of the dividend-price ratio together with the lagged real dividend growth rate, to show significant predictive power on stock returns.

These studies are typically performed by regressing the excess returns on a constant and a lagged variable. The conventional T-test is then used to make inference concerning predictability. However, most of these studies are based on the presumption of the stationarity of the predictors, where the T-statistic is approximately normally distributed in large samples. Unfortunately, this is not the case in the presence of highly persistent variables. Even when the predictors are stationary, asymptotic critical values are not a good approximation for those obtained in finite-sample distributions. In the presence of highly persistent predictors, the innovations are greatly correlated with the returns, and thus, the T-statistic has a non-standard distribution which leads to the over-rejection of the null hypothesis of orthogonality [see. Elliott and Stock (1994), Mankiw and Shapiro (1986), Stambaugh (1999) and Campbell and Yogo (2006)].

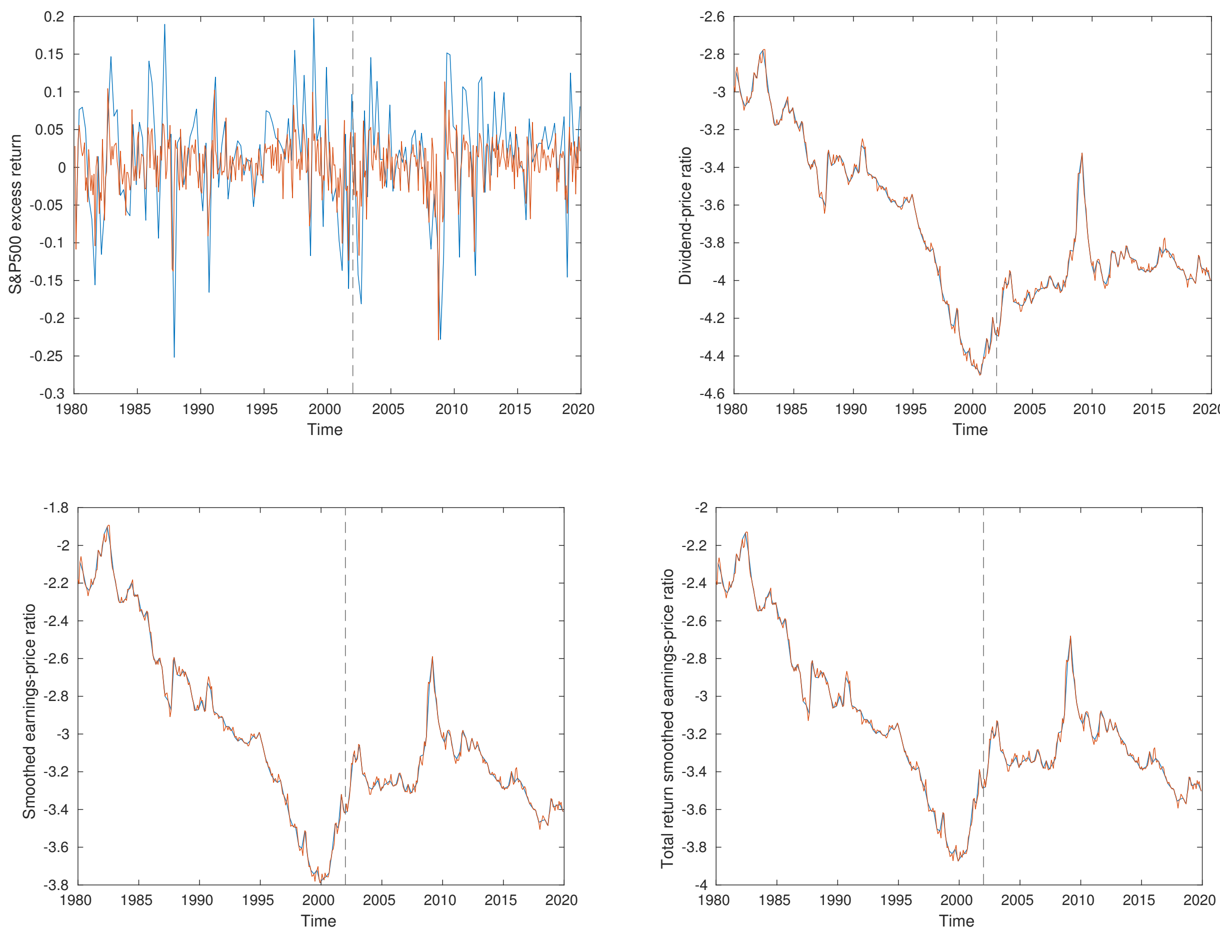
Most studies address the issue of persistency by making inference based on more accurate approximations of the finite-sample distribution of the test-statistic. This is accomplished either by relying on exact finite-sample theory under the assumption of normality [see. Evans and Savin (1981, 1984) and Stambaugh (1999)] or local-to-unity asymptotics [see Elliott and Stock (1994), Campbell and Yogo (2006) and Torous et al. (2004)]. More recently Taamouti et al. (2014) confirm the predictability power of the valuation ratios using monthly data, in a nonparametric and model-free copula-based Granger causality framework.

In this Section, we use our exact 10% SS-POS-based test to make inference and compare the predictive power of the valuation ratios (dividend-price ratio, smoothed earnings-price ratio, and total return smoothed earnings-price ratio) on stock market returns. The smoothed earnings-price ratio is proposed by Campbell and Shiller (1988, 2001) upon observing numerous spikes in the plot of the earnings-price ratio that had not been observed in the dividend-price ratio. The spikes were explained to be caused by recessions, which temporarily suppress corporate earnings. The latter measure is the ratio of the ten-year moving average of real earnings to current real prices and is said to possess better forecasting powers. Furthermore, the total return smoothed earnings-price ratio is recently incorporated in forecasting, as a consequence of the changes in corporate payout policy documented by Bunn et al. (2014) and Jivraj and Shiller (2017). Share repurchases (as opposed to dividends) have become the dominant approach for distributing cash to shareholders in the U.S. which may impact the smoothed earnings-price ratio through changes in growth of earnings per share. The total return smoothed earnings-price ratio corrects for this bias by reinvesting the dividends into the price index, such that the earnings per share is appropriately scaled.

1.6.1.1 Data description

Our data consists of monthly and quarterly observations of the aggregate S&P500 composite index for the period spanning from March 1980 to December 2019 for a total of 480 trading months or 160 trading quarters. We consider the logarithmic returns on the S&P500 in excess of the 30-day and 90-day T-bill rate. The valuation ratios under consideration are: dividend-price ratio, smoothed earnings-price ratio, and total return smoothed earnings-price ratio. The nominal monthly and quarterly prices of the value-weighted S&P500 composite index, as well as the corresponding dividends and earnings are obtained from a database provided on Robert Shiller's website. The 30-day and 90-day Treasury bill returns, on the other hand, have been retrieved from the Center for Research in Security Prices (CRSP).

Figure 1.9: Monthly and quarterly S&P500 excess stock returns, dividend-price, smoothed earnings-price and total return smoothed earnings-price ratios.



Note: The data spans from March 1980 to December 2019 for a total of 480 trading months and 160 trading quarters respectively. The red and the blue lines in turn correspond to the quarterly and monthly samples. To assess the predictability power of the valuation ratios, we further consider two sub-periods separated by the dashed line: one spanning from March 1980 to January 2002 and another in the period of January 2002 to January 2019.

in different At first glance figure 1.9 suggests that the predictors under consideration are highly persistent and potentially non-stationary. This visual assessment is confirmed in table 1.2, which presents the test statistics for the augmented Dickey-Fuller test (ADF hereafter) for all the time series. Evidently, for the full sample and the two sub-periods we fail to reject the null hypothesis of nonstationarity. The testing procedure entails estimating and testing the model in its most general form using more deterministic components than the hypothesized DGP (i.e. including both an intercept and a trend), and following Phillips and Perron (1988) sequential testing strategy thereafter, eliminating the unnecessary nuisance parameters in the process. At each stage, if the null hypothesis of orthogonality is rejected, we conclude that the model is correctly specified and that the process is stationary. Otherwise, the test is performed on a more restricted model. This procedure is continued until we arrive at the most basic form of the model (with no intercept or a trend), or until the null hypothesis of unit root is rejected. As it is evident, all valuation ratios reject the null hypothesis of non-stationarity at the 5% level.

1.6.1.2 Predictability results

The projection technique based on the proposed 10% SS-POS test is used to build simultaneous confidence sets for the parameters of the regressions of the excess returns against the dividend-price ratio, smoothed earnings-price ratio of Campbell and Shiller (1988) and the total return smoothed earnings-price ratio of Bunn et al. (2014) and Jivraj and Shiller (2017) respectively. The results for different sub-periods and the full sample are reported in table 1.3. As explained in Section 1.4, each simultaneous confidence set is obtained by collecting all pairs of (β_0, β_1) that are not rejected using our 10% SS-POS test. Thus, a grid search is applied over an appropriate range[1] and 95% level confidence sets are constructed by retaining all the pairs (β_0, β_1) that are not rejected by the 10% SS-POS test. Alternatively, the simulated annealing algorithm can be used to solve the optimization problem (1.15) for each parameter β_i.

The 95% confidence intervals for the parameters β_0 and β_1 contain zero for the regressions of the excess returns against all the predictors using the T-test based on White (1980) for all periods in our study. However, using the 10% SS-POS based test, there is evidence of predictability in quarterly data in favor of all predictors for the period spanning from January 2002 to January 2019. Our findings are in line with those of Campbell and Yogo (2006) who do not find any evidence of predictability in favor of any of the predictors in the period spanning from 1952-2002.

[1]See Section 1.4.1.

Table 1.2: Results of the ADF test on the real and nominal time-series using the general-to-specific sequential testing procedure

Series	Obs.	Predictor	p	$\delta+\mu$	μ	None
Panel A: 1980-2002						
Monthly	264	$r_t^m - r_t^f$	1	-10.959^{***}	$--$	$--$
		d/p_t	2	-2.217	-0.657	2.026
		$e/p_t^{'}$	2	-2.248	-1.171	1.721
		$e/p_t^{''}$	2	-2.160	-1.376	1.544
Quarterly	88	$r_t^m - r_t^f$	0	-9.026^{***}	$--$	$--$
		d/p_t	0	-2.209	-0.777	1.830
		$e/p_t^{'}$	0	-1.816	-1.210	1.576
		$e/p_t^{''}$	0	-1.669	-1.400	1.391
Panel B: 2002-2019						
Monthly	215	$r_t^m - r_t^f$	0	-11.369^{***}	$--$	$--$
		d/p_t	1	-2.853	-2.983^{**}	$--$
		$e/p_t^{'}$	1	-2.317	-1.938	-0.027
		$e/p_t^{''}$	1	-2.389	-1.935	0.009
Quarterly	72	$r_t^m - r_t^f$	0	-7.513^{***}	$--$	$--$
		d/p_t	1	-3.261^{*}	-3.278^{**}	$--$
		$e/p_t^{'}$	0	-2.374	-1.915	-0.095
		$e/p_t^{''}$	0	-2.448	-1.901	-0.057
Panel C: 1980-2019						
Monthly	479	$r_t^m - r_t^f$	1	-14.347^{***}	$--$	$--$
		d/p_t	2	-1.861	-2.104	0.935
		$e/p_t^{'}$	2	-1.802	-2.042	1.136
		$e/p_t^{''}$	2	-1.965	-2.161	1.056
Quarterly	160	$r_t^m - r_t^f$	0	-11.848^{***}	$--$	$--$
		d/p_t	0	-1.762	-2.051	0.876
		$e/p_t^{'}$	0	-1.732	-1.995	1.084
		$e/p_t^{''}$	0	-1.897	-2.114	0.998

Note: This table reports the results of the ADF test on the time-series in the predictive regression model. The appraoch involves using the general-to-specific sequential testing procedure to test the null hypothesis of non-stationarity, where the general form of the model is:

$$\Delta x_t = \rho x_{t-1} + \sum_{i=1}^{p-1} \psi_i \Delta x_{t-i} + \mu + \delta t + u_t \quad u_t \sim IID(0, \sigma^2) \qquad .$$

The corresponding test statistics are reported in turn for the general form of the model (including the trend δ and intercept c), the more restrictive form constituting only of an intercept c, and the case where neither the trend nor the intercept are present. The variables are defined as follows: $r_t^m - t_t^f$ are the excess logarithmic stock returns, d/p_t is the dividend-price ratio, e/p_t' is the smoothed earnings-price ratio and e/p_t'' is the total return smoothed earnings-price ratio respectively. The statistics with three asterisks (***), two asterisks (**) and one asterisk (*) are significant at the 1%, 5%. and the 10% levels respectively.

Table 1.3: Predictability results for the dividend-price, earnings-price and the smoothed earnings-price ratios

Series	Predictor	$\hat{\beta}$	95% confidence interval	
			10% SS-POST	WT-test
Panel A: 1980-2002				
Monthly	d/p_t	0.002	$[-0.024, 0.036]$	$[-0.008, 0.011]$
	$e/p_t^{'}$	-0.001	$[-0.044, 0.046]$	$[-0.009, 0.008]$
	$e/p_t^{''}$	-0.001	$[-0.052, 0.049]$	$[-0.010, 0.010]$
Quarterly	d/p_t	0.009	$[-0.104, 0.106]$	$[-0.028, 0.047]$
	$e/p_t^{'}$	0.003	$[-0.116, 0.104]$	$[-0.029, 0.036]$
	$e/p_t^{''}$	0.004	$[-0.126, 0.104]$	$[-0.033, 0.040]$
Panel B: 2002-2019				
Monthly	d/p_t	0.019	$[-0.220, 0.330]$	$[-0.015, 0.053]$
	$e/p_t^{'}$	0.012	$[-0.079, 0.191]$	$[-0.018, 0.042]$
	$e/p_t^{''}$	0.010	$[-0.080, 0.180]$	$[-0.021, 0.040]$
Quarterly	d/p_t	0.119	$[\mathbf{0.159}, \mathbf{0.899}]$	$[-0.001, 0.238]$
	$e/p_t^{'}$	0.089	$[\mathbf{0.042}, \mathbf{0.632}]$	$[-0.018, 0.197]$
	$e/p_t^{''}$	0.084	$[\mathbf{0.058}, \mathbf{0.697}]$	$[-0.026, 0.194]$
Panel C: 1980-2019				
Monthly	d/p_t	0.002	$[-0.041, 0.069]$	$[-0.006, 0.010]$
	$e/p_t^{'}$	0.0003	$[-0.021, 0.049]$	$[-0.007, 0.007]$
	$e/p_t^{''}$	0.0001	$[-0.039, 0.061]$	$[-0.008, 0.008]$
Quarterly	d/p_t	0.136	$[-0.094, 0.146]$	$[-0.017, 0.044]$
	$e/p_t^{'}$	0.008	$[-0.099, 0.121]$	$[-0.020, 0.036]$
	$e/p_t^{''}$	0.009	$[-0.113, 0.147]$	$[-0.023, 0.041]$

Note: This table presents the coefficient estimates, as well as the 95% confidence intervals for the variables considered in our study, by inverting the proposed 10% SS-POS-based tests and the T-test based on White (1980) variance correction. The alternatives for the 10% SS-POS tests are obtained by running OLS regressions of the excess returns against the dividend-price, smoothed earnings-price and the total return smoothed earnings-price ratios. The regressions assume the form

$$r_t^m - r_t^f = \beta_0 + \beta_1 x_{t-1} + \varepsilon_t \qquad (1.19)$$

where r_t is the ex-post excess returns and x_{t-1} is the ex-ante predictor. The projection-based 95% confidence intervals for the 10% SS-POS tests are obtained by testing $H_0(\beta^*) : \beta = \beta^*$ on a grid for $\beta^* = (\beta_0^*, \beta_1^*)$, where the grid dimension is found by solving the optimization problem (1.15) for each parameter β_0 and β_1 using the simulated annealing algorithm, and consequently equally dividing each interval and finding their Cartesian product. The intervals in bold do not contain the value of zero and imply significance at the 5% level.

1.7 Conclusion

In this chapter, we proposed simple point-optimal sign-based tests for inference in linear and non-linear predictive regression models in the presence of stochastic (or fixed) regressors. One motivation of the paper is to build valid (control the size whatever the sample size) tests for linear and non-linear predictability of stock returns. The most popular predictors of stock returns (e.g. dividend-price ratio, earning-price ratio, etc.) are known to be persistent with residuals that are correlated with the shock in the stock returns. This makes the classical predictability tests not valid, especially when the sample size is small or moderate. In addition, the proposed sign-based tests are exact, distribution-free, and robust against heteroskedasticity of unknown form and allow for serial (non-linear). Additionally, they may be inverted to build confidence regions for the parameters of the regression function. Since the point-optimal sign tests depend on the alternative hypothesis, an adaptive approach based on the split-sample technique was suggested in order to choose the appropriate alternative that controls the size and maximizes the power.

We presented a Monte Carlo study to assess the performance of the proposed "quasi"-point-optimal sign test by comparing its size and power to those of certain existing tests which are supposed to be robust against heteroskedasticity. We considered different DGPs to illustrate different contexts that one can encounter in practice. The results show that the 10% split-sample point-optimal sign test is more powerful than the T-test, Campbell and Dufour (1995) sign-based test, and the T-test based on White (1980) variance correction.

Finally, the proposed tests were used to assess the predictive power of some financial predictors, such as the dividend-price ratio, earnings-price ratio and the smoothed earnings-price ratio of Campbell and Shiller (1988, 2001) on the annualized monthly excess stock returns. Our study suggests predictability in favor of all the predictors for the quarterly data in the period spanning from 2002 to 2009. which is consistent with the findings of Campbell and Yogo (2006), Our findings are in line with those of Campbell and Yogo (2006) who do not find any evidence of predictability in favor of any of the predictors in the period spanning from 1952-2002.

1.8 Appendix: Proofs

Proof of Theorem 1. From Assumption (1.2), the following two equalities are derived

$$P[\varepsilon_t \geq 0 \mid X] = \mathbb{E}\left(P[\varepsilon_t \geq 0 \mid \boldsymbol{\varepsilon_{t-1}}, X]\right) = \frac{1}{2}$$

with

$$\boldsymbol{\varepsilon}_0 = \{\emptyset\}, \quad \boldsymbol{\varepsilon}_{t-1} = \{\varepsilon_1, \cdots, \varepsilon_{t-1}\}, \quad \text{for} \quad t \geq 2$$

and

$$P[\varepsilon_t \geq 0 \mid \underline{\mathrm{S}}^{\varepsilon}_{t-1}, X] = P[\varepsilon_t \geq 0 \mid \boldsymbol{\varepsilon_{t-1}}, X] = \frac{1}{2},$$

with

$$\underline{\mathrm{S}}^{\varepsilon}_0 = \{\emptyset\}\,, \quad \underline{\mathrm{S}}^{\varepsilon}_{t-1} = \{s(\varepsilon_1) = s_1, ..., s(\varepsilon_{t-1}) = s_{t-1}\}\,, \text{ for } t \geq 2,$$

We define the vector of signs $U(n) = (s(y_1), \cdots, s(y_n))'$, where $s(y_t) = \mathbb{1}_{\mathbb{R}^+ \cup 0}\{y_t\}$. Thus, the likelihood function of the sample in terms of signs under the null hypothesis is

$$\begin{aligned} L(U(n), 0) &= P[s(y_1) = s_1, \cdots, s(y_n) = s_n \mid X] \\ &= P[s(\varepsilon_1) = s_1, \cdots, s(\varepsilon_n) = s_n \mid X] \\ &= \prod_{t=1}^{n} P[\varepsilon_t \geq 0 \mid \boldsymbol{\varepsilon}_{t-1}, X]^{s(\varepsilon_t)} \left(1 - P[\varepsilon_t \geq 0 \mid \boldsymbol{\varepsilon}_{t-1}, X]\right)^{1-s(\varepsilon_t)} \\ &= \prod_{t=1}^{n} \left(\frac{1}{2}\right)^{s(\varepsilon_t)} \left(1 - \frac{1}{2}\right)^{1-s(\varepsilon_t)} \\ &= \left(\frac{1}{2}\right)^{n} \end{aligned}$$

Hence, it can be concluded that conditional on X and under the null hypothesis of orthogonality $s(y_1), \cdots, s(y_n) \overset{i.i.d}{\sim} Bi(1, 0.5)$. ■

Proof of Proposition 1. The likelihood function of sample in terms of signs $s(y_1), ..., s(y_n)$

$$L\left(U(n), \beta\right) = P\left[s(y_1) = s_1, ..., s(y_n) = s_n \mid X\right] = \prod_{t=1}^{n} P\left(s(y_t) = s_t \mid \underline{\mathrm{S}}_{t-1}, X\right),$$

for

$$\underset{\sim}{S}_0 = \{\emptyset\}\,, \quad \underset{\sim}{S}_{t-1} = \{s(y_1) = s_1, ..., s(y_{t-1}) = s_{t-1}\}\,, \text{ for } t \geq 2,$$

and

$$P\left[s(y_1) = s_1 \mid \underset{\sim}{S}_0, X\right] = P\left[s(y_1) = s_1 \mid X\right],$$

where each s_i, for $1 \leq t \leq n$, takes two possible values 0 and 1. According to model (1.1) and assumption (1.2), under the null hypothesis the signs $s(y_1), \cdots, s(y_n)$ are i.i.d according to $Bi(1, 0.5)$,

$$P\left[s(y_t) = 1 \mid X\right] = P\left[s(y_t) = 0 \mid X\right] = \frac{1}{2}, \text{ for } t = 1, ..., n.$$

Consequently, under H_0

$$L_0\left(U(n), 0\right) = \prod_{t=1}^{n} P\left[s(y_t) = s_t \mid X\right] = \left(\frac{1}{2}\right)^n$$

and under H_1 we have

$$L_1\left(U(n), \beta_1\right) = \prod_{t=1}^{n} P\left[s(y_t) = s_t \mid \underset{\sim}{S}_{t-1}, X\right]$$

where now, for $t = 1, ..., n$,

$$y_t = \beta_1' x_{t-1} + \varepsilon_t$$

The log-likelihood ratio is given by

$$\ln\left\{\frac{L_1\left(U(n), \beta_1\right)}{L_0\left(U(n), 0\right)}\right\} = \sum_{t=1}^{n} \ln\left\{P\left[s(y_t) = s_t \mid \underset{\sim}{S}_{t-1}, X\right]\right\} - \text{n} \ln\left\{\frac{1}{2}\right\}.$$

According to Neyman-Pearson lemma [see e.g. Lehmann (1959), page 65], the best test to test H_0 against H_1, based on $s(y_1), ..., s(y_n)$, rejects H_0 when

$$\ln\left\{\frac{L_1\left(U(n), \beta_1\right)}{L_0\left(U(n), 0\right)}\right\} \geq c$$

or when

$$\sum_{t=1}^{n} \ln\left\{P\left[s(y_t) = s_t \mid \underset{\sim}{S}_{t-1}, X\right]\right\} \geq c,$$

The critical value, say c, is given by the smallest constant c such that

$$P\left(\ln\left\{\frac{L_1\left(U(n),\beta_1\right)}{L_0\left(U(n),0\right)}\right\} > c \mid H_0\right) \leq \alpha.$$

Notice that, for $t = 1, ...n$,

$$P\left[s(y_t) = s_t \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right] = P\left[y_t \geq 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]^{s(y_t)} P\left[y_t < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]^{(1-s(y_t))}, \text{ for } t = 1, ...n. \tag{1.20}$$

From (1.20), we have

$$\begin{aligned}
\ln\left\{\prod_{t=1}^{n} P\left[s(y_t) = s_t \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]\right\} &= \ln\left\{\prod_{t=1}^{n} P[y_t \geq 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X]^{s(y_t)} P[y_i < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X]^{(1-s(y_t))}\right\} \\
&= \sum_{t=1}^{n} s(y_t) \ln\left\{P\left[y_t \geq 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]\right\} \\
&\quad + \sum_{t=1}^{n} (1 - s(y_t)) \ln\left\{P[y_t < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X]\right\}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{t=1}^{n} s(y_t) \ln\left\{P\left[y_t \geq 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]\right\} + \sum_{t=1}^{n} \ln\left\{P[y_t < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X]\right\} \\
&\quad - \sum_{t=1}^{n} s(y_t) \ln\left\{P[y_t < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X]\right\} \\
&= \sum_{t=1}^{n} s(y_t) \ln\left\{\frac{P\left[y_t \geq 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}{P\left[y_t < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}\right\} + \sum_{t=1}^{n} \ln\left\{P[y_t < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X]\right\}
\end{aligned}$$

Thus, the best test to test H_0 against H_1, based on $s(y_1), ..., s(y_n)$, rejects H_0 when

$$\ln\left\{\frac{L_1\left(U(n),\beta_1\right)}{L_0\left(U(n),0\right)}\right\} = \sum_{t=1}^{n} s(y_t) \ln\left\{\frac{P\left[y_t \geq 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}{P\left[y_t < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}\right\} + \sum_{t=1}^{n} \ln\left\{P[y_t < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X]\right\} - \mathrm{n} \ln\left\{\frac{1}{2}\right\} \geq c$$

or when

$$\ln\left\{\frac{L_1\left(U(n),\beta_1\right)}{L_0\left(U(n),0\right)}\right\} = \sum_{t=1}^{n} s(y_t) \ln\left\{\frac{P\left[y_t \geq 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}{P\left[y_t < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}\right\} \geq c_1(\beta_1)$$

where the critical value $c_1(\beta_1)$ is chosen so that

$$P\left[S_n(\beta_1) > c_1(\beta_1) \mid H_0\right] \leq \alpha$$

α is an arbitrary significance level. ■

Proof of Assumption 1.

Let $y_1, \cdots, y_t$ be linearly explained by a vector variable x_t

$$y_t = \beta_1' x_{t-1} + \varepsilon_t$$

and suppose y_t follows a Markov process of order one, such that

$$y_t, y_{t-1} \mid X \sim N\left(\underbrace{\begin{bmatrix} \beta' x_{t-1} \\ \beta' x_{t-2} \end{bmatrix}}_{\boldsymbol{\mu}}, \underbrace{\begin{bmatrix} \sigma^2_{\varepsilon_t} & \sigma_{\varepsilon_t \varepsilon_{t-1}} \\ \sigma_{\varepsilon_{t-1}\varepsilon_t} & \sigma^2_{\varepsilon_{t-1}} \end{bmatrix}}_{\Sigma}\right), \quad t = 2, \cdots, n.$$

where it is assumed, without loss of generality, that ε_t and ε_{t-1} are serially correlated with $\sigma_{\varepsilon_{t-1}\varepsilon_t} \neq 0$. Then the signs $s(y_1), \cdots, s(y_t)$ are Bernoulli variables with conditional joint distributions fully determined by $P_{s(y_t)|X}$, $P_{s(y_{t-1})|X}$, and either $P_{s(y_t)|s(y_{t-1}),X}$ or $P_{s(y_t),s(y_{t-1})|X}$, where

$$\begin{aligned} P_{s(y_t)|s(y_{t-1}),X} &:= P[s(y_t) = 1 \mid s(y_{t-1}) = 1, X] \\ P_{s(y_t),s(y_{t-1})|X} &:= P[s(y_t) = 1, s(y_{t-1}) = 1 \mid X], \end{aligned}$$

which may alternatively be expressed as

$$P[s(y_t) = 1 \mid s(y_{t-1}) = 1, X] = P[y_t \geq 0 \mid y_{t-1} \geq 0 \mid X] = P[\varepsilon_t \geq -\beta' x_{t-1} \mid \varepsilon_{t-1} \geq -\beta' x_{t-2}, X]$$
$$P[s(y_t) = 1, s(y_{t-1}) = 1 \mid X] = P[y_t \geq 0, y_{t-1} \geq 0 \mid X] = P[\varepsilon_t \geq -\beta' x_{t-1}, \varepsilon_{t-1} \geq -\beta' x_{t-2} \mid X].$$

As the dependence in the pair-wise probabilities is determined by the covariance matrix Σ, with $\sigma_{\varepsilon_t,\varepsilon_{t-1} \neq 0}$, this in turn implies that the signs $s(y_t), s(y_{t-1})$ for $t = 1, \cdots, n$ are dependent and follow a Markov process of order one. These findings are extended to the case where the signs exhibit non-linear serial dependence.

■

Proof of Corollary 1. From test statistic $S_n(\beta_1)$ in Proposition 1 and under assumption **A1**, we have:

$$\begin{aligned}
\tilde{S}_n(\beta_1) &= \sum_{t=1}^{n} s(y_t) \ln\left\{\frac{P\left[y_t \geq 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}{P\left[y_t < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}\right\} \\
&= \sum_{t=1}^{n} s(y_t) \left\{\ln\left\{P\left[y_t \geq 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]\right\} - \ln\left\{P\left[y_t < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]\right\}\right\} \\
&= \sum_{t=1}^{n} s(y_t) \left\{\begin{array}{c} \ln\left\{P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]^{s(y_{t-1})} P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]^{1-s(y_{t-1})}\right\} \\ -\ln\left\{P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right]^{s(y_{t-1})} P\left[y_t < 0 \mid y_{t-1} < 0, X\right]^{1-s(y_{t-1})}\right\} \end{array}\right\} \\
&= \sum_{t=1}^{n} s(y_t) \left\{\begin{array}{c} s(y_{t-1}) \ln\left\{P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]\right\} + (1 - s(y_{t-1})) \ln\left\{P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]\right\} \\ -s(y_{t-1}) \ln\left\{P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right]\right\} - (1 - s(y_{t-1})) \ln\left\{P\left[y_t < 0 \mid y_{t-1} < 0, X\right]\right\} \end{array}\right\}
\end{aligned}$$

Observe that:

$$\begin{aligned}
\ln\left\{\frac{P\left[y_t \geq 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}{P\left[y_t < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}\right\} &= \ln\left\{P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]^{s(y_{t-1})} P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]^{1-s(y_{t-1})}\right\} \\
&\quad - \ln\left\{P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right]^{s(y_{t-1})} P\left[y_t < 0 \mid y_{t-1} < 0, X\right]^{1-s(y_{t-1})}\right\} \\
&= s(y_{t-1}) \ln\left\{P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]\right\} \\
&\quad + (1 - s(y_{t-1})) \ln\left\{P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]\right\} \\
&\quad - s(y_{t-1}) \ln\left\{P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right]\right\} \\
&\quad - (1 - s(y_{t-1})) \ln\left\{P\left[y_t < 0 \mid y_{t-1} < 0, X\right]\right\}
\end{aligned}$$

$$\begin{aligned}
&= s(y_{t-1}) \ln\left\{P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]\right\} + \ln\left\{P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]\right\} \\
&\quad - s(y_{t-1}) \ln\left\{P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]\right\} - s(y_{t-1}) \ln\left\{P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right]\right\} \\
&\quad - \ln\left\{P\left[y_t < 0 \mid y_{t-1} < 0, X\right]\right\} + s(y_{t-1}) \ln\left\{P\left[y_t < 0 \mid y_{t-1} < 0, X\right]\right\}
\end{aligned}$$

$$= \quad s(y_{t-1})\left\{\ln\left\{\frac{P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]}{P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right]}\right\} - \ln\left\{\frac{P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]}{P\left[y_t < 0 \mid y_{t-1} < 0, X\right]}\right\}\right\}$$
$$+\ln\left\{\frac{P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]}{P\left[y_t < 0 \mid y_{t-1} < 0, X\right]}\right\}$$

Hence,

$$\begin{aligned}
\tilde{S}_n(\beta_1) &= \sum_{t=1}^{n} s(y_t)\ln\left\{\frac{P\left[y_t \geq 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}{P\left[y_i < 0 \mid \underset{\sim}{\mathrm{S}}_{t-1}, X\right]}\right\} \\
&= \sum_{t=1}^{n} s(y_t)\left\{\begin{array}{c} s(y_{t-1})\left\{\ln\left\{\frac{P[y_t \geq 0|y_{t-1} \geq 0,X]}{P[y_t<0|y_{t-1}\geq 0,X]}\right\} - \ln\left\{\frac{P[y_t\geq 0|y_{t-1}<0,X]}{P[y_t<0|y_{t-1}<0,X]}\right\}\right\} \\ +\ln\left\{\frac{P[y_t\geq 0|y_{t-1}<0,X]}{P[y_t<0|y_{t-1}<0,X]}\right\} \end{array}\right\} \\
&= \sum_{t=1}^{n} s(y_t)\ln\left\{\frac{P\left[y_t \geq 0 \mid y_t < 0, X\right]}{P\left[y_t < 0 \mid y_t < 0, X\right]}\right\} + \sum_{t=1}^{n} s(y_t)s(y_{t-1})\left\{\begin{array}{c} \ln\left\{\frac{P[y_t\geq 0|y_{t-1}\geq 0,X]}{P[y_t<0|y_{t-1}\geq 0,X]}\right\} \\ -\ln\left\{\frac{P[y_t\geq 0|y_{t-1}<0,X]}{P[y_t<0|y_{t-1}<0,X]}\right\} \end{array}\right\} \\
&= \sum_{t=1}^{n} a_t\; s(y_t) + \sum_{t=1}^{n} b_t\; s(y_t)s(y_{t-1})
\end{aligned}$$

where

$$\tilde{a}_1 = \ln\left\{\frac{P\left[y_1 \geq 0 \mid X\right]}{P\left[y_1 < 0 \mid X\right]}\right\} = \ln\left\{\frac{1 - P\left[\varepsilon_1 < -\beta_1' x_0 \mid X\right]}{P\left[\varepsilon_1 < -\beta_1' x_0 \mid X\right]}\right\}$$

$$\tilde{b}_1 = \ln\left\{\frac{P\left[y_1 \geq 0 \mid X\right]}{P\left[y_1 < 0 \mid X\right]}\right\} - \ln\left\{\frac{P\left[y_1 \geq 0 \mid X\right]}{P\left[y_1 < 0 \mid X\right]}\right\} = 0$$

and for $t = 2, ..., n$

$$a_t \quad = \quad \ln\left\{\frac{P\left[y_t \geq 0 \mid y_t < 0, X\right]}{P\left[y_t < 0 \mid y_t < 0, X\right]}\right\},$$

$$b_t \quad = \quad \ln\left\{\frac{P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]}{P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right]}\right\} - \ln\left\{\frac{P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right]}{P\left[y_t < 0 \mid y_{t-1} < 0, X\right]}\right\}.$$

Observe that:

$$
\begin{aligned}
P\left[y_t \geq 0 \mid y_{t-1} < 0, X\right] &= 1 - P\left[y_t < 0 \mid y_{t-1} < 0, X\right] \\
&= 1 - \frac{P\left[y_t < 0,\ y_{t-1} < 0 \mid X\right]}{P\left[y_{t-1} < 0 \mid X\right]} \\
&= 1 - \frac{P\left[\varepsilon_t < -\beta_1' x_{t-1},\ \varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}{P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}, \\
P\left[y_t < 0 \mid y_{t-1} < 0, X\right] &= \frac{P\left[y_t < 0,\ y_{t-1} < 0 \mid X\right]}{P\left[y_{t-1} < 0 \mid X\right]} \\
&= \frac{P\left[\varepsilon_t < -\beta_1' x_{t-1},\ \varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}{P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]} \\
P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right] &= 1 - P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right] \\
&= 1 - \frac{P\left[y_t < 0,\ y_{t-1} \geq 0 \mid X\right]}{P\left[\ y_{t-1} \geq 0 \mid X\right]} \\
&= 1 - \frac{P\left[y_t < 0 \mid X\right]}{P\left[\ y_{t-1} \geq 0 \mid X\right]}\left(P\left[\ y_{t-1} \geq 0 \mid y_t < 0, X\right]\right) \\
&= 1 - \frac{P\left[y_t < 0 \mid X\right]}{P\left[\ y_{t-1} \geq 0 \mid X\right]}\left(1 - P\left[\ y_{t-1} < 0 \mid y_t < 0, X\right]\right) \\
&= 1 - \left(\frac{P\left[y_t < 0 \mid X\right]}{P\left[\ y_{t-1} \geq 0 \mid X\right]} - \frac{P\left[\ y_{t-1} < 0,\ y_t < 0 \mid X\right]}{P\left[\ y_{t-1} \geq 0 \mid X\right]}\right) \\
&= 1 - \left(\frac{P\left[y_t < 0 \mid X\right]}{1 - P\left[y_{t-1} < 0 \mid X\right]} - \frac{P\left[\ y_{t-1} < 0,\ y_t < 0 \mid X\right]}{1 - P\left[y_{t-1} < 0 \mid X\right]}\right) \\
&= 1 - \left[\frac{P\left[\varepsilon_t < -\beta_1' x_{t-1} \mid X\right]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]} - \frac{P\left[\ \varepsilon_{t-1} < -\beta_1' x_{t-2},\ \varepsilon_t < -\beta_1' x_{t-1} \mid X\right]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' x_{t-2} \mid X\right]}\right)
\end{aligned}
$$

$$
\begin{aligned}
P\left[y_t < 0 \mid y_{t-1} \geq 0, X\right] &= \frac{P\left[y_t < 0,\ y_{t-1} \geq 0 \mid X\right]}{P\left[y_{t-1} \geq 0 \mid X\right]} \\
&= \frac{P\left[y_{t-1} \geq 0 \mid\ y_t < 0, X\right] P\left[y_t < 0 \mid X\right]}{P\left[y_{t-1} \geq 0 \mid X\right]} \\
&= \frac{P\left[y_t < 0 \mid X\right]}{P\left[y_{t-1} \geq 0 \mid X\right]}\left(1 - P\left[y_{t-1} < 0 \mid\ y_t < 0, X\right]\right) \\
&= \frac{P\left[y_t < 0 \mid X\right]}{1 - P\left[y_t < 0 \mid X\right]} - \frac{P\left[y_{t-1} < 0,\ y_t < 0 \mid X\right]}{1 - P\left[y_t < 0 \mid X\right]} \\
&= 1 - P\left[y_t \geq 0 \mid y_{t-1} \geq 0, X\right]
\end{aligned}
$$

We also have:

$$\begin{aligned}
\tilde{a}_1 &= \ln\left\{\frac{P\left[y_1 \geq 0 \mid X\right]}{P\left[y_1 < 0 \mid X\right]}\right\} \\
&= \ln\left\{\frac{1 - P\left[y_1 < 0 \mid X\right]}{P\left[y_1 < 0 \mid X\right]}\right\} \\
&= \ln\left\{\frac{1 - P\left[\varepsilon_1 < -\beta_1' x_0 \mid X\right]}{P\left[\varepsilon_1 < -\beta_1' x_0 \mid X\right]}\right\}
\end{aligned}$$

$$\tilde{b}_1 = \ln\left\{\frac{P\left[y_1 \geq 0 \mid X\right]}{P\left[y_1 < 0 \mid X\right]}\right\} - \ln\left\{\frac{P\left[y_1 \geq 0 \mid X\right]}{P\left[y_1 < 0 \mid X\right]}\right\} = 0$$

■

The simulated annealing algorithm. In this Section, we adapt the global numerical optimization search algorithm of Goffe et al. (1994) to obtain valid confidence intervals for the parameters β_j in (1.8) by solving the problem

$$\min_{\beta \in \mathbb{R}^k} \ \beta_j \ \ s.c. \ \widehat{SN}_n(\beta_0 \mid \beta_1) < \tilde{c}_1(\beta_0, \beta_1), \quad \max_{\beta \in \mathbb{R}^k} \ \beta_j \ \ s.c. \ \widehat{SN}_n(\beta_0 \mid \beta_1) < \tilde{c}_1(\beta_0, \beta_1)$$

where the critical value $\tilde{c}(\beta_0, \beta_1)$ at level α, is computed using B replications of the statistic $\widehat{SN}_n^{(i)}(\beta_0 \mid \beta_1)$ under the null hypothesis and finding its $(1-\alpha)$ quantile. In what follows, we only consider the maximization problem, noting that the minimization algorithm is almost identical:

(I) **set** $C \leftarrow 0.25$ as the speed of adjustment of V;

(II) **set** $\beta_0 \leftarrow [\underbrace{0 \ \ \ldots \ \ 0}_{k}]$ starting vector of parameters;

(III) **set** $V \leftarrow [\underbrace{0.5 \ \ \ldots \ \ 0.5}_{\dim(\beta_0)}]$, which must cover the entire range of interest in parameter β_0;

(IV) **set** $\varepsilon \leftarrow 0.01$ as the convergence criteria;

(V) **set** $r_T \leftarrow 0.75$ as the temperature reduction factor;

(VI) **set** $T_0 \leftarrow 50$ as the initial temperature;

(VII) **set** $N_\varepsilon \leftarrow 10$ as the number of times through function before termination;

(VIII) **set** $N_S \leftarrow 20$ as the number of times through function before V adjustment;

(IX) **set** $N_T \leftarrow 20$ as the number of times through N_S loops before T reduction;

The algorithm is as follows

1. Calculate the alternative hypothesis β_1 using the 10% split-sample technique;
2. Let $\beta_{opt} \leftarrow \beta_0$ and $f_{opt} \leftarrow \beta_{opt}$;
3. **do until convergence**
4. **do** N_T **times**
5. **do** N_S **times**
6. **for** $j = 1, ..., \dim(\beta_0)$ **do**
7. Allocate a $\dim(\beta_0) \times 1$ vector to β';
8. $\beta_0^{j'} \leftarrow \beta_0^j + U \times V_j$ where U is uniformly distributed on $[-1, 1]$;
9. Let $f' \leftarrow \beta_0'$;
10. Evaluate the test statistic $\widehat{SN}_n(\beta_0' \mid \beta_1)$;
11. Simulate the distribution of the test statistic $\widehat{SN}_n(\beta_0' \mid \beta_1)$ under β_0' and find the critical value $\tilde{c}_1(\beta_0', \beta_1)$;
12. **if** $\widehat{SN}_n(\beta_0' \mid \beta_1) < \tilde{c}_1(\beta_0', \beta_1)$ **then**
13. **if** $f' \leq f$ **then**
14. $p \leftarrow \exp((f' - f)/T_0)$;
15. $p' \leftarrow rand \sim U[0, 1]$;
16. **if** $p > p'$ **then**
17. $\beta_0 \leftarrow \beta_0'$, $f \leftarrow f'$, $\beta_{opt} \leftarrow \beta_0'$, $f_{opt} \leftarrow f'$;
18. **end if**

end if

end if

if $f' > f$ **then**

$\beta_0 \leftarrow \beta_0'$, $f \leftarrow f'$;

end if

if $f' > f_{opt}$

$\beta_0 \leftarrow \beta_0'$, $f \leftarrow f'$, $\beta_{opt} \leftarrow \beta_0'$, $f_{opt} \leftarrow f'$;

end if

end if

end for

end do

Adjust V such that half of all trials are accepted;

end do

if $\Delta f_{opt} < \varepsilon$ last N_ε iterations $|f - f'| < \varepsilon$ **then**

Report β_{opt}, f_{opt} & V;

stop

else

$\beta_0 \leftarrow \beta_{opt}$ (start on current best optimum);

$T \leftarrow r_T \times T$ (reduce T);

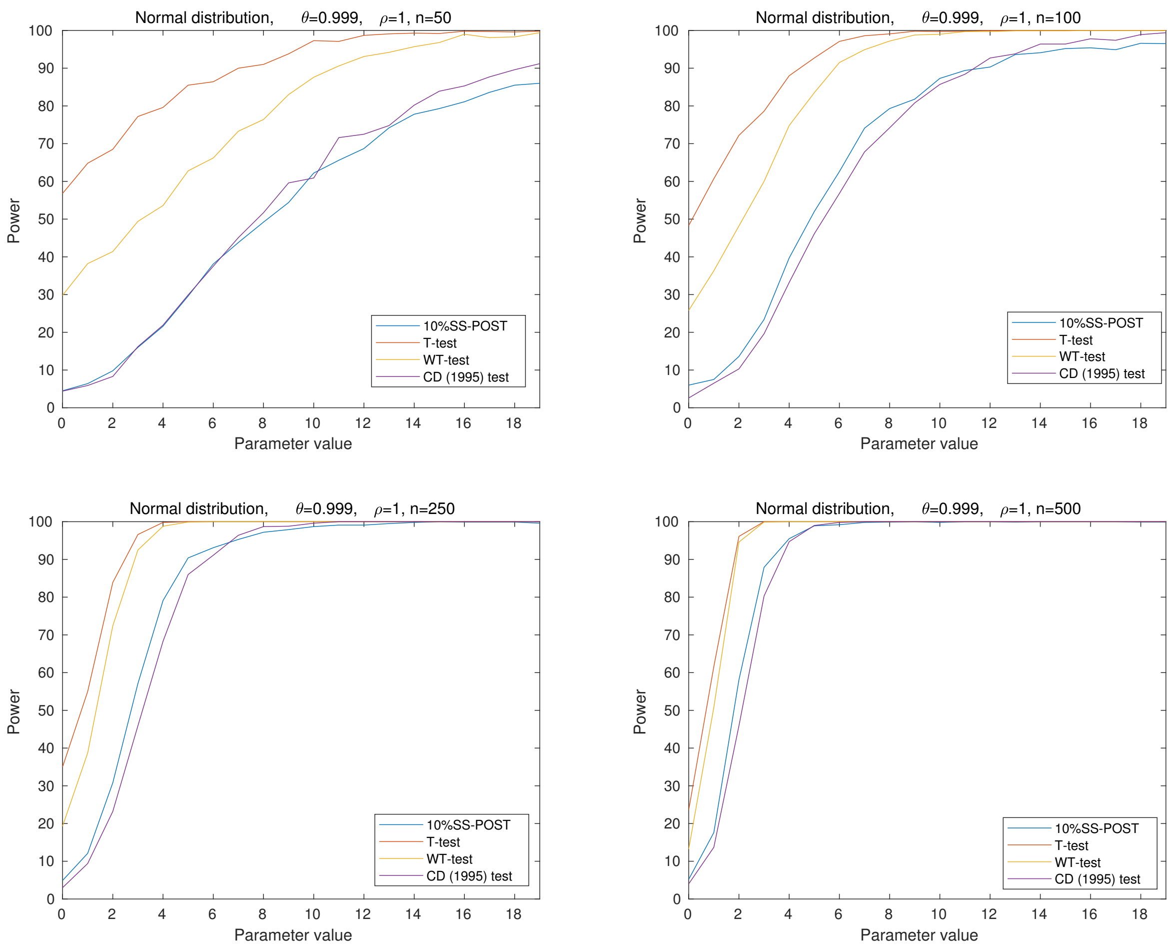
end if

39. **continue**

■

Additional simulations.

Figure 1.10: Power comparisons: different tests. Normal distributions with contemporaneous correlation of $\rho = 1$, in (1.18) and local-to-unity autoregression parameter $\theta = 0.999$, in (1.17) for different sample sizes.



Note: These figures compare the power function of the 10% SS-POS test with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

■

Chapter 2

Pair copula constructions of point-optimal sign-based tests for predictive linear and non-linear regressions

2.1 Introduction

The disturbances of regressions often exhibit non-normal distributions and heteroskedasiticty of unknown form, in the presence of which parametric tests perform poorly in terms of size control and power in finite samples. In an extensive simulations exercise, Dufour and Taamouti (2010a) show that the heteroskedasticity and autocorrelation corrected tests developed by White (1980) (more commonly referred to as "HAC"procedures) are plagued with low power when the errors follow GARCH structures or there is a break in the variance. To address these issues, Dufour and Taamouti (2010a) propose point-optimal sign-based inference to test whether the conditional median of a response variable is zero against a linear regression alternative, where this procedure is further extended to non-linear models. These tests are constructed by considering fixed regressors and error terms that are independent with zero median conditional on the explanatory variables.

The first chapter of this thesis is an extension of the sign tests developed by Dufour and Taamouti (2010a), in which we propose exact point-optimal sign-based tests (POS-based tests hereafter) to test for predictability in the presence of highly persistent stochastic regressors for both linear and non-linear models. However, in order to obtain *feasible* POS-based test statistics, we had to impose a Markovian assumption on the sign process. The aim of this paper is to relax the Markovianity assumption for the POS-based tests. This can be achieved using the models introduced by Panagiotelis et al. (2012) for multivariate discrete data based on pair copula constructions (PCC hereafter). These models would allow us to build feasible test statistics that are robust against heavy-tailed and asymmetric distributions, provided that the errors have zero median conditional on their own past and the explanatory variables, without the necessity to impose any additional (and potentially restrictive) assumptions.

As noted earlier, when the predictors follow a local-to-unity autoregression, there is a high degree of contemporaneous correlation between the errors in the regressors and the disturbances of the predictive regresssion. In such situation, least-squares based T-type tests possess a non-standard distribution and inference using asymptotic critical values is no longer valid [see Mankiw and Shapiro (1986) and Stambaugh (1999) among others]. As the POS-based type tests such as those introduced by Dufour and Taamouti (2010a) are randomized tests with a randomized distribution under the null hypothesis [see Pratt and Gibbons (2012)], the said procedures do not suffer from the issues encountered by T-type statistics in finite samples. Therefore, by relaxing the independence assumption on the error terms and by allowing the disturbances to exhibit serial (non-linear) dependence, the POS-based tests are easily extended to a predictive regression framework. The POS-based tests are shown to be robust against non-standard distributions and heteroskedasticity of unknown form and to have the highest power among parametric and nonparametric tests that are supposed to be robust against heteroskedasticity. Moreover, as in Dufour and Taamouti (2010a) they can be inverted to produce a confidence region for the vector (sub-vector) of parameters.

Although, the literature surrounding sign-based and sign-ranked inference is vast [see Taamouti (2015) and Boldin et al. (1997) among others], the focus of the POS-based tests constructed by Dufour and Taamouti (2010a) is to maximize power at a nominated point in the alternative parameter space. As such, the power of the POS-based test is close to that of the power envelope - i.e. maximum attainable power for a given testing problem [see King (1987)]. Therefore, the POS-based tests in Dufour and Taamouti (2010a) and those developed in this chapter using the pair copula construction of discrete data (PCC-POS-based tests hereafter) are Neyman-Pearson type tests based on the signs, and as in Dufour and Taamouti (2010a) a practical problem concerns finding an alternative at which the power of the PCC-POS-based tests is close to that of the power envelope. By conducting an intensive simulations exercise, Dufour and Taamouti (2010a) find that when 10% of the sample is used to estimate the alternative and the remaining portion is used to calculate the test-statistic, the power of the POS-based test traces out the power envelope. Our simulations results using the 10% split-sample PCC-POS-based tests confirm these findings.

Many studies have developed distribution-free sign and sign-ranked statistics that are exact and robust against different forms of heteroskedasticity. These range from the procedures proposed for bivariate regressions [see. Campbell and Dufour (1991, 1995, 1997) and Luger (2003) among others], to those for multivariate regressions [see Dufour and Taamouti (2010a)]. In the context of dependent data, Coudin and Dufour (2009) extend the procedures proposed by Boldin et al. (1997) to further consider serial dependence, as well as discrete distributions. The work in this chapter, as well as the first chapter of this thesis fall within the latter category (i.e. sign-based testing procedures for dependent data), and they are particularly motivated by the regressions capturing the predictability of stock returns. Predictors of stock returns, such as earnings-price and dividend price ratios often possess relatively static numerators and contain the non-stationary price series in their denominator; hence, as noted earlier, these predictors are shown to be highly persistent, with innovations that are correlated with the residuals of the predictive regressions, which lead to invalid inference [see Mankiw and Shapiro (1986) and Stambaugh (1985, 1999)].

Due to the non-linear nature of the signs, there is inherent uncertainty regarding the structure of sign dependence. Therefore, it is important to consider the entire dependence structure of the signs. One approach for computing the joint distribution of the signs $s(y_1), ..., s(y_n)$, where $s(y_i) = \mathbb{1}_{\mathbb{R}^+ \cup \{0\}}\{y_i\}$, entails taking advantage of copula functions [see Sklar (1959)], which express the joint distribution of the signs in terms of the i) marginal distributions of the individual signs;

and ii) copula models capturing the dependence of the n signs. As the signs are discrete, the likelihood function of the POS-based tests under the alternative hypothesis can then be calculated using rectangle probabilities and in turn estimated using copulae with closed analytical form. However, this approach would not yield feasible test statistics, as the number of *multivariate* copulae that need to be evaluated increase at an exponential rate as the sample size n increases. As a result of this curse of dimensionality, the literature surrounding calculating probability mass functions (p.m.f hereafter) using discrete data is limited to low-dimensional data and copulae that are fast to calculate [see Nikoloulopoulos and Karlis (2008, 2009) and Li and Wong (2010)].

To propose feasible test statistics, we build POS-based tests in the context of stochastic regressors for linear and non-linear models, using a discrete analogue of the vine PCCs proposed by Panagiotelis et al. (2012). The likelihood function of the signs under the alternative hypothesis can be decomposed as a vine PCC under a set of conditions that are later outlined in the paper. The most important advantage of the latter method is that for a sample of size n, only $2n(n-1)$ *bivariate* copula evaluations are required, as opposed to 2^n *multivariate* copula evaluations using the rectangle probabilities approach. Another advantage of the vine PCC methodology is that model selection techniques can be used to identify the conditional independence in the process of signs in order to create more parsimonious PCC models.

An issue that needs considerable attention is whether estimating $2n(n-1)$ parameters for evaluating the bivariate copulae in the PCC-POS-based tests is feasible as n tends to infinity. Even when more parsimonius PCC models are selected, this approach is only feasible in finite samples. However, in a strict stationarity framework, the parameters are invariant to time shifts and as such this number drastically reduces to $n-1$ parameter estimates, which may further reduce to as small as one parameter in the case of truncated PCC models. A Monte Carlo study reveals that pair copula constructions of POS-based tests are valid. Furthermore, under *most* distributional assumptions, they possess the maximum power among tests that are intended to be robust against non-standard distributions and heteroskedasticity of unknown form.

The outline of the paper is as follows: in Section 2.2, we motivate the use of the discrete analogue of the vine PCC for building POS-based tests. In Section 2.3, we outline the conditions under which vine PCCs can be implemented and we also discuss the choice of the PCC model. We then propose PCC-POS-based tests for linear and non-linear models. In secion 2.4, we discuss the estimation approach implemented for the vine PCCs. In Section 2.5, we discuss the choice of the alternative hypothesis for computing the PCC-POS-based test statistic. In Section 2.6, we discuss the problem of finding a confidence set for a vector (subvector) of parameters using the projection techniques. In Section 2.7, we assess the performance of the proposed tests in terms of size and power. Finally, in Section 2.8 we conclude the findings of the paper.

2.2 Framework

Consider a stochastic process $Z = \{Z_t = (y_t, x_t') : \Omega \to \mathbb{R}^{(k+1)} : t = 0, 1, \cdots\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. Suppose that y_t can linearly be explained by a vector variable x_t

$$y_t = \beta' x_{t-1} + \varepsilon_t, \quad t = 1, ..., n, \tag{2.1}$$

where x_{t-1} is an $(k+1)\times 1$ vector of stochastic explanatory variables, say $x_{t-1} = [1, x_{1,t-1}, ..., x_{k,t-1}]'$, $\beta \in \mathbb{R}^{(k+1)}$ is an unknown vector of parameters with $\beta = [\beta_0, \beta_1, ..., \beta_k]'$ and

$$\varepsilon_t \mid X \sim F_t(. \mid X)$$

where $F_t(. \mid X)$ is an unknown conditional distribution function and X=$[x_0, \cdots, x_{n-1}]'$ is an $n \times (k+1)$ matrix.

As in the first chapter, we follow Coudin and Dufour (2009) by considering the median as an alternative measure of central tendency. This implies imposing a median-based analogue of the martingale difference sequence (MDS) on the error process - namely we suppose that ε_t is a strict conditional mediangale

$$P[\varepsilon_t > 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = P[\varepsilon_t < 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = \frac{1}{2}, \tag{2.2}$$

with

$$\boldsymbol{\varepsilon}_0 = \{\emptyset\}, \quad \boldsymbol{\varepsilon}_{t-1} = \{\varepsilon_1, \cdots, \varepsilon_{t-1}\}, \quad \text{for} \quad t \geq 2$$

Note (2.2) entails that $\varepsilon_t \mid X$ has no mass at zero for all t, which is only true if $\varepsilon_t \mid X$ is a continuous variable. Model (2.1) in conjunction with assumption (2.2) allows the error terms to possess asymmetric, heteroskedastic and serially dependent distributions, so long as the conditional medians are zero. Assumption 1.2 allows for many dependent schemes, such as those of the form $\varepsilon_1 = \sigma_1(x_1, \cdots, x_{t-2})\epsilon_1$, $\varepsilon_t = \sigma_1(x_1, \cdots, x_{t-2}, \varepsilon_1, \cdots, \varepsilon_{t-1})\epsilon_t$,$t = 2, \cdots, n$, where $\epsilon_1, \cdots, \epsilon_n$ are independent with a zero median. In time-series context this includes models such as ARCH, GARCH or stochastic volatility with non-Gaussian errors. Furthermore, in the mediangale framework the disturbances need not be second order stationary.

Suppose, we wish to test the null hypothesis

$$H_0 : \beta = 0, \tag{2.3}$$

against the alternative

$$H_1 : \beta = \beta_1. \tag{2.4}$$

Define the vector of signs as follows

$$U(n) = (s(y_1), ..., s(y_n))',$$

where for $t = 1, ..., n$

$$s(y_t) = \begin{cases} 1, \; if \;\; y_t \geq 0 \\ 0, \; if \;\; y_t < 0 \end{cases} .$$

We consider Neyman-Pearson type test based on the signs. Thus, to build POS-based tests for testing the null hypothesis (2.3) against the alternative (2.4), we first define the likelihood function of the sample in terms of signs $s(y_1), ..., s(y_n)$

$$L(U(n), \beta) = P[s(y_1) = s_1, ..., s(y_n) = s_n \mid X] = \prod_{t=1}^{n} P\left[s(y_t) = s_t \mid \mathrm{S}_{t-1} = \underline{\mathrm{s}}_{t-1}, X\right], \tag{2.5}$$

with

$$\underset{\sim}{S}_0 = \{\emptyset\}, \quad \underset{\sim}{S}_{t-1} = \{s(y_1), ..., s(y_{t-1})\}, \quad \text{for} \quad t \geq 2,$$

and

$$P[s(y_1) = s_1 \mid \underset{\sim}{S}_0 = \underset{\sim}{s}_0, X] = P[s(y_1) = s_1 \mid X],$$

where each s_t for $1 \leq t \leq n$ takes two possible values of 0 and 1. Under model (2.1) and assumption (2.2), the variables $s(\varepsilon_1), \cdots, s(\varepsilon_n)$ and in turn $s(y_1), \cdots, s(y_n)$ are i.i.d conditional on X, according to the distribution

$$P[s(\varepsilon_1) = 1 \mid X] = P[s(\varepsilon_1) = 0 \mid X] = \frac{1}{2}, \quad t = 1, \cdots, n$$

This results holds true iff for any combination of $t = 1, \cdots, n$ there is a permutation $\pi : i \to j$ such that the mediangale assumption holds for j. Then the signs $s(\varepsilon_1), \cdots, s(\varepsilon_n)$ are i.i.d. [see Theorem 2]. Therefore, under the null hypothesis we have

$$P[s(y_t) = 1 \mid X] = P[s(y_t) = 0 \mid X] = \frac{1}{2}, \quad t = 1, ..., n. \tag{2.6}$$

Consequently, under the null hypothesis of orthogonality, the log-likelihood function is given by

$$L_0(U(n), 0) = \prod_{t=1}^{n} P[s(y_t) = s_t \mid X] = \left(\frac{1}{2}\right)^n.$$

On the other hand, under the alternative we have

$$L_1(U(n), \beta_1) = \prod_{t=1}^{n} P[s(y_t) = s_t \mid \underset{\sim}{S}_{t-1} = \underset{\sim}{s}_{t-1}, X],$$

where now for $t = 1, .., n$

$$y_t = \beta_1' x_{t-1} + \varepsilon_t.$$

In the first chapter, we considered optimal sign-based tests (in the Neyman-Pearson sense), which maximize power under the constraint $P[\text{Reject } H_0 \mid H_0] \leq \alpha$, where α is an arbitrary significance level [see Lehmann and Romano (2006)]. Let H_0 and H_1 be defined by (2.3) and (2.4) respectively.

Then under the assumptions (2.1) and (2.2), the log-likelihood ratio

$$SL_n(\beta_1) = \ln\left\{\frac{L_1(U(n), \beta_1)}{L_0(U(n), 0)}\right\} > c, \tag{2.7}$$

is most powerful for testing H_0 against H_1 among level α tests based on the signs $(s(y_1), ..., s(y_n))'$, where c is the smallest constant such that

$$P[SL_n(\beta_1) > c \mid H_0] \leq \alpha,$$

where α is an arbitrary significance level and where $L_0(U(n), 0)$ is the likelihood function under H_0.

In chapter one, it has been shown that in the presence of stochastic regressors, the test statistic requires the calculation of $P[y_t \geq 0 \mid \underset{\sim}{S}_{t-1} = \underset{\sim}{s}_{t-1}, X]$ and $P[y_t < 0 \mid \underset{\sim}{S}_{t-1} = \underset{\sim}{s}_{t-1}, X]$. The latter is not easy to compute, as it involves the distribution of the joint process of signs $s(y_1), ..., s(y_n)$, conditional on X which is unknown. Therefore, to obtain feasible test statistics, we made an assumption that the sign process $\{s(y_t)\}_{t=0}^{\infty}$ follows a Markov process of finite order; in our study, we considered a Markov process of order one. However, it may be important to capture the dependence structure of the entire process.

An approach by which we may consider the entire dependence structure of the vector of signs is to take advantage of copulae. The Theorem of Sklar (1959) states that there exists a copula C such that

$$F(s_1, ..., s_n \mid X) = C(F_1(s_1 \mid X), ..., F_n(s_n \mid X)), \tag{2.8}$$

where F is a conditional joint cumulative distribution function (CDF hereafter) of the vector of signs $\underset{\sim}{S} = (s(y_1), ..., s(y_n))'$ with conditional marginal distribution functions F_j for $j = 1, 2, ..., n$. Copula C is unique for continuous variables, but for discrete variables it is unique only on the set

$$\text{Range}(F_1) \times ... \times \text{Range}(F_n),$$

which is the Cartesian product of the ranges of the marginal distribution functions. To illustrate an example of non-uniqueness in the discrete case, let us consider a sample of two discrete binary variables, say $s(y_1)$ and $s(y_2)$, with corresponding marginal distribution functions F_1 and F_2. We know that $F_j \sim \text{Bernoulli}(p_j)$ for $j = 1, 2$, such that

$$F_j(s_j \mid X) = \begin{cases} 0, & \text{for} \quad s_j < 0 \\ 1 - p_j, & \text{for} \quad 0 \leq s_j < 1 \\ 1, & \text{for} \quad s_j \geq 1 \end{cases} \tag{2.9}$$

Thus, $\text{Range}(F_1) = \{0, 1-p_1, 1\}$ and $\text{Range}(F_2) = \{0, 1-p_2, 1\}$, with the copula only being unique for $C(1-p_1, 1-p_2)$, noting that $C(0, 1-p_j) = 0$ and $C(1, 1-p_j) = 1-p_j$ for $j = 1, 2$. However, this non-uniqueness does not preclude the use of parametric copulae for modelling discrete data [see. Joe (1997), Song et al. (2009)]. Considering this bivariate example, the p.m.f can be expressed in terms of rectangle probabilities,

$$\begin{aligned} P[s(y_1) = s_1, s(y_2) = s_2 \mid X] &= P[s_1 - 1 < s(y_1) \leq s_1, s_2 - 1 < s(y_2) \leq s_2 \mid X] \\ &= F(s_1, s_2 \mid X) - F(s_1 - 1, s_2 \mid X) \\ &\quad -F(s_1, s_2 - 1 \mid X) + F(s_1 - 1, s_2 - 1 \mid X) \end{aligned}$$

and in turn in terms of copulae as follows

$$\begin{aligned} P[s(y_1) = s_1, s(y_2) = s_2 \mid X] &= F(s_1, s_2 \mid X) - F(s_1 - 1, s_2 \mid X) \\ &\quad -F(s_1, s_2 - 1 \mid X) + F(s_1 - 1, s_2 - 1 \mid X) \\ &= C(F_1(s_1 \mid X), F_2(s_2 \mid X)) - C(F_1(s_1 - 1 \mid X), F_2(s_2 \mid X)) \\ &\quad -C(F_1(s_1 \mid X), F_2(s_2 - 1 \mid X)) + C(F_1(s_1 - 1 \mid X), F_2(s_2 - 1 \mid X)), \end{aligned}$$

which implies that the n-variate likelihood function (2.5) can be expressed in terms of 2^n finite differences

$$\begin{aligned}P[s(y_1)=s_1,...,s(y_n)=s_n \mid X] &= \sum_{i_1=0,1} ... \sum_{i_n=0,1} (-1)^{i_1+...+i_n} P[s(y_1) \leq s_1 - i_1, ..., s(y_n) \leq s_n - i_n \mid X] \\ &= \sum_{i_1=0,1} ... \sum_{i_n=0,1} (-1)^{i_1+...+i_n} C(F_1(s_1 - i_1 \mid X), ..., F_n(s_n - i_n \mid X)).\end{aligned}$$

Evidently, the calculation of likelihood function (2.5) using this approach would require 2^n *multivariate* copula evaluations, which is not computationally feasible. However, by employing the vine PCC introduced later in the paper, we will show that this number can be reduced to only $2n(n-1)$ *bivariate* copula evaluations. The latter method provides us with flexibility, since any multivariate discrete distribution can be decomposed as a vine PCC under a set of conditions that are discussed in the following Section.

2.3 Pair copula constructions of point-optimal tests

In this Section, we expand on the first chapter by deriving POS-based tests in the context of linear and non-linear regression models based on vine PCC decomposition. Following a structure similar to Dufour and Taamouti (2010a), we first consider the problem of testing whether the conditional median of a vector of observations is zero against a linear regression alternative. We further consider the conditions under which the likelihood function under the alternative can be decomposed as a vine PCC, and as such, choose an appropriate vine model. These results are later generalized to test whether the coefficients of a possibly non-linear median regression function have a given value against an alternative non-linear median regression.

2.3.1 Testing independence (zero coefficients) hypothesis in linear regressions

Consider the problem of testing the null hypothesis (2.3) against the alternative (2.4), using the test statistic (2.7) and given the assumptions (2.1) and (2.2). As it was shown in Section 2.2, under the alternative hypothesis the likelihood function can be expressed as

$$L_1(U(n), \beta_1) = \prod_{t=1}^{n} P\left[s(y_t) = s_t \mid \underset{\sim}{\mathrm{S}}_{t-1} = \underset{\sim}{\mathrm{s}}_{t-1}, X\right]. \tag{2.10}$$

Let $s(y_j)$ be a scalar element of $\underset{\sim}{\mathrm{S}}_{t-1}$, with $\underset{\sim}{\mathrm{S}}_{t-1}^{\backslash j} = \underset{\sim}{\mathrm{S}}_{t-1} \backslash s(y_j)$ such that

$$\underset{\sim}{\mathrm{S}}_{t-1}^{\backslash j} = \{s(y_1), s(y_2), ..., s(y_{j-1}), s(y_{j+1}), ..., s(y_{t-1})\}$$

and $s(y_t) \notin \underset{\sim}{\mathrm{S}}_{t-1}$. By choosing a single element of $\underset{\sim}{\mathrm{S}}_{t-1}$, say $s(y_j)$, we would have

$$\begin{aligned} P\left[s(y_t) = s_t \mid \underset{\sim}{\mathrm{S}}_{t-1} = \underset{\sim}{\mathrm{s}}_{t-1}, X\right] &= \frac{P[s(y_t) = s_t, s(y_j) = s_j \mid \underset{\sim}{\mathrm{S}}_{t-1}^{\backslash j} = \underset{\sim}{\mathrm{s}}_{t-1}^{\backslash j}, X]}{P[s(y_j) = s_j \mid \underset{\sim}{\mathrm{S}}_{t-1}^{\backslash j} = \underset{\sim}{\mathrm{s}}_{t-1}^{\backslash j}, X]} \\ &= \sum_{k_t=0,1} \sum_{k_j=0,1} (-1)^{k_t+k_j} \times \\ &\quad \left\{ P[s(y_t) \leq s_t - k_t, s(y_j) \leq s_j - k_j \mid \underset{\sim}{\mathrm{S}}_{t-1}^{\backslash j} = \underset{\sim}{\mathrm{s}}_{t-1}^{\backslash j}, X] \right\} \\ &\quad / P[s(y_j) = s_j \mid \underset{\sim}{\mathrm{S}}_{t-1}^{\backslash j} = \underset{\sim}{\mathrm{s}}_{t-1}^{\backslash j}, X], \end{aligned} \tag{2.11}$$

where the bivariate conditional probability in (2.11) can be expressed in terms of copulae as follows

$$\begin{aligned} P[s(y_t) = s_t \mid \underset{\sim}{\mathrm{S}}_{t-1} = \underset{\sim}{\mathrm{s}}_{t-1}, X] = \sum_{k_t=0,1} \sum_{k_j=0,1} (-1)^{k_t+k_j} \Bigg\{ \\ C_{s(y_t), s(y_j) \mid \underset{\sim}{\mathrm{S}}_{t-1}^{\backslash j}} \left(F_{s(y_t) \mid \underset{\sim}{\mathrm{S}}_{t-1}^{\backslash j}}(s_t - k_t \mid \underset{\sim}{\mathrm{s}}_{t-1}^{\backslash j}, X), F_{s(y_j) \mid \underset{\sim}{\mathrm{S}}_{t-1}^{\backslash j}}(s_j - k_j \mid \underset{\sim}{\mathrm{s}}_{t-1}^{\backslash j}, X) \right) \Bigg\} \\ / P[s(y_j) = s_j \mid \underset{\sim}{\mathrm{S}}_{t-1}^{\backslash j} = \underset{\sim}{\mathrm{s}}_{t-1}^{\backslash j}, X]. \end{aligned} \tag{2.12}$$

Further, let $\underline{\mathrm{S}}_{t-1}^{\backslash i,j} = \underline{\mathrm{S}}_{t-1}^{\backslash j} \backslash s(y_i)$, such that $s(y_i)$ is a scalar element of $\underline{\mathrm{S}}_{t-1}^{\backslash j}$. Then the arguments $F_{s(y_t)|\underline{\mathrm{S}}_{t-1}^{\backslash j}}$ and $F_{s(y_j)|\underline{\mathrm{S}}_{t-1}^{\backslash j}}$ in copula expression (2.12) can be expressed by the general form

$$\begin{aligned}
&F_{s(y_t)|s(y_i),\underline{\mathrm{S}}_{t-1}^{\backslash i,j}}(s_t - k_t \mid s_i, \underline{\mathrm{s}}_{t-1}^{\backslash i,j}, X) = \\
&\Big\{C_{s(y_t),s(y_i)|\underline{\mathrm{S}}_{t-1}^{\backslash i,j}}\Big(F_{s(y_t)|\underline{\mathrm{S}}_{t-1}^{\backslash i,j}}(s_t - k_t \mid \underline{\mathrm{s}}_{t-1}^{\backslash i,j}, X), F_{s(y_i)|\underline{\mathrm{S}}_{t-1}^{\backslash i,j}}(s(y_i) \mid \underline{\mathrm{s}}_{t-1}^{\backslash i,j}, X)\Big) - \\
&\quad C_{s(y_t),s(y_i)|\underline{\mathrm{S}}_{t-1}^{\backslash i,j}}\Big(F_{s(y_t)|\underline{\mathrm{S}}_{t-1}^{\backslash i,j}}(s_t - k_t \mid \underline{\mathrm{s}}_{t-1}^{\backslash i,j}, X), F_{s(y_i)|\underline{\mathrm{S}}_{t-1}^{\backslash i,j}}(s(y_i) - 1 \mid \underline{\mathrm{s}}_{t-1}^{\backslash i,j}, X)\Big)\Big\} \\
&/P[s(y_i) = s_i \mid \underline{\mathrm{S}}_{t-1}^{\backslash i,j} = \underline{\mathrm{s}}_{t-1}^{\backslash i,j}, X].
\end{aligned} \tag{2.13}$$

Thus, decomposition (2.12), and in turn (2.13) can be applied recursively to the elements of the likelihood function (2.5), such that it is expressed in terms of bivariate copulae. Let $\underline{\mathrm{S}}_{t-1} = \{s(y_1), ..., s(y_{t-1})\}$ be the variables that $s(y_t)$ for $t = 2, ..., n$ is conditioned on. We follow Joe (2014), by letting $\underline{\sigma}_{t-1} = \{\sigma(1,t), ..., \sigma(t-1,t)\}$ be a permutation of $\underline{\mathrm{S}}_{t-1}$, such that $s(y_t)$ is paired sequentially, first to $\sigma(1,t)$, then $\sigma(2,t)$ and finally $\sigma(t-1,t)$, where in the r^{th} step $(2 \le r \le t-1)$, $\sigma(r,t)$ is paired to t conditional on $\sigma(1,t), ..., \sigma(r-1,t)$. For $n \le 3$ (i.e. $t = 2, 3$) there are only three possible permutations with $\underline{\sigma}_1 = \{s(y_1)\}$ for $t = 2$, and $\underline{\sigma}_2 = \{s(y_1), s(y_2)\}$, as well as $\underline{\sigma}_2 = (s(y_2), s(y_1))$ for $t = 3$ respectively. Therefore, under assumptions (2.1) and (2.2), and with $n \le 3$, let H_0 and H_1 be defined by (2.3) - (2.4), then the Neyman-Pearson type test-statistic based on the signs $(s(y_1), ..., s(y_n))'$ can be expressed as

$$\begin{aligned}
SL_n(\beta_1) = \ln P\left[s(y_1) = s_1 \mid X\right] \quad &+ \quad \sum_{t=2}^{n} \ln \Delta_{s_t^-}^{s_t^+} \Delta_{s_{t-1}^-}^{s_{t-1}^+} C_{t,t-1|t-2} \\
&- \sum_{t=2}^{n} \ln P[s(y_{t-1}) = s_{t-1} \mid \underline{\mathrm{S}}_{t-2} = \underline{\mathrm{s}}_{t-2}, X] - n \ln \left\{\frac{1}{2}\right\},
\end{aligned}$$

for $t = 2, ..., n$, where

$$\begin{aligned}
\Delta_{s_t^-}^{s_t^+} \Delta_{s_{t-1}^-}^{s_{t-1}^+} C_{t,t-1|t-2} \quad = \quad & \sum_{k_t=0,1} \sum_{k_{t-1}=0,1} (-1)^{k_t + k_{t-1}} \\
& \times \left(C_{s(y_t),s(y_{t-1})|\underline{\mathrm{S}}_{t-2}}\left(F_{s(y_t)|\underline{\mathrm{S}}_{t-2}}(s_t - k_t \mid \underline{\mathrm{s}}_{t-2}, X), F_{s(y_{t-1})|\underline{\mathrm{S}}_{t-2}}(s_{t-1} - k_{t-1} \mid \underline{\mathrm{s}}_{t-2}, X)\right)\right)
\end{aligned}$$

and such that

$$\ln P\left[s(y_1) = s_1 \mid \underset{\sim}{\mathrm{S}}_0 = \underset{\sim}{\mathrm{s}}_0, X\right] = s(y_1) \ln \left\{ \frac{P[y_1 \geq 0 \mid X]}{P[y_1 < 0 \mid X]} \right\} + \ln P[y_1 < 0 \mid X].$$

where $X = [x_0, ..., x_{n-1}]'$ is a $n \times (k+1)$ matrix of stochastic explanatory variables.

Proof: *See Appendix.*

For $n > 3$, the permutations $\underset{\sim}{\sigma}_{t-1}$ are dependent on the choice of the permutations at stages $3, ..., t-1$. Therefore, an issue that requires considerable attention is whether there *exists* a decomposition such as the one considered in the earlier example, for $n > 3$. Furthermore, the likelihood function expressed in terms of bivariate copulae using by recursively using (2.12) and (2.13), assumes that a single copula is specified for each conditional bivariate distribution $F_{s(y_t),s(y_j)|\underset{\sim}{\mathrm{S}}_{t-1}^{\setminus j}}$ in decomposition (2.10) over all possible values of $\underset{\sim}{\mathrm{S}}_{t-1}^{\setminus j}$, which implies that the copula is unique for the Cartesian product of the ranges of conditional CDFs $F_{s(y_t)|\underset{\sim}{\mathrm{S}}_{t-1}^{\setminus j}}$ and $F_{s(y_j)|\underset{\sim}{\mathrm{S}}_{t-1}^{\setminus j}}$. Therefore, the decomposition must be such that each conditional bivariate distribution in the said vine has a constant conditional copula [see Panagiotelis et al. (2012)]. For a constant conditional copula to exist, the following conditions outlined by Panagiotelis et al. (2012) must be satisfied.

Existence of constant conditional copula: *Consider the bivariate distribution function* $F_{s(y_t),s(y_j)|\underset{\sim}{S}_{t-1}^{\setminus j}}$. *We say that a copula* $C = C_{s(y_t),s(y_j)|\underset{\sim}{S}_{t-1}^{\setminus j}}$ *is constant over all possible values of* $\underset{\sim}{S}_{t-1}^{\setminus j}$ *if*

$$\sum_{m=0,1} \sum_{n=0,1} (-1)^{m+n} C(a_{k-m}, b_{l-n}) \geq 0, \quad \forall k, l \in \{1,2\} \times \{1,2\},$$

where $a_0 < a_1 < a_2$ *and* $b_0 < b_1 < b_2$*, are the distinct points corresponding to the ranges of the conditional Bernoulli CDFs* $F_{s(y_t)|\underset{\sim}{S}_{t-1}^{\setminus j}}$ *and* $F_{s(y_j)|\underset{\sim}{S}_{t-1}^{\setminus j}}$ *respectively, such that* $a_0 = b_0 = 0$ *and* $a_2 = b_2 = 1$*, and where further, the following constraints are satisfied:*

$$C_{s(y_t),s(y_j)|\underset{\sim}{\mathrm{S}}_{t-1}^{\setminus j}}\left(a_{s(y_t)|\underset{\sim}{\mathrm{S}}_{t-1}^{\setminus j}}, b_{s(y_j)|\underset{\sim}{\mathrm{S}}_{t-1}^{\setminus j}}\right) = P[s(y_t) \leq s_t, s(y_j) \leq s_j \mid \underset{\sim}{\mathrm{S}}_{t-1}^{\setminus j} = \underset{\sim}{\mathrm{s}}_{t-1}^{\setminus j}, X],$$

$$C_{s(y_t),s(y_j)|\underset{\sim}{\mathrm{S}}_{t-1}^{\setminus j}}\left(1, b_{s(y_j)|\underset{\sim}{\mathrm{S}}_{t-1}^{\setminus j}}\right) = b_{s(y_j)|\underset{\sim}{\mathrm{S}}_{t-1}^{\setminus j}}, \quad C_{s(y_t),s(y_j)|\underset{\sim}{\mathrm{S}}_{t-1}^{\setminus j}}\left(a_{s(y_t)|\underset{\sim}{\mathrm{S}}_{t-1}^{\setminus j}}, 1\right) = a_{s(y_t)|\underset{\sim}{\mathrm{S}}_{t-1}^{\setminus j}},$$

where $a_{s(y_t)|\underset{\sim}{S}_{t-1}^{\setminus j}} := P[s(y_t) \leq s_t \mid \underset{\sim}{S}_{t-1}^{\setminus j} = \underset{\sim}{s}_{t-1}^{\setminus j}, X]$ *and* $b_{s(y_j)|\underset{\sim}{S}_{t-1}^{\setminus j}} := P[s(y_j) \leq s_j \mid \underset{\sim}{S}_{t-1}^{\setminus j} = \underset{\sim}{S}_{t-1}^{\setminus j}, X]$.

To satisfy the above conditions, the vine decomposition must be such that the strength of the dependence of the conditional bivariate distribution does not vary much across different values of the conditioning set [see Panagiotelis et al. (2012)]. As we are dealing with time-series data, the D-vine decomposition yields a constant dependence structure over different values of $\underset{\sim}{S}_{t-1}^{\setminus j}$, and is thus, the most appropriate and intuitive choice for the decomposition of the likelihood function (2.10).

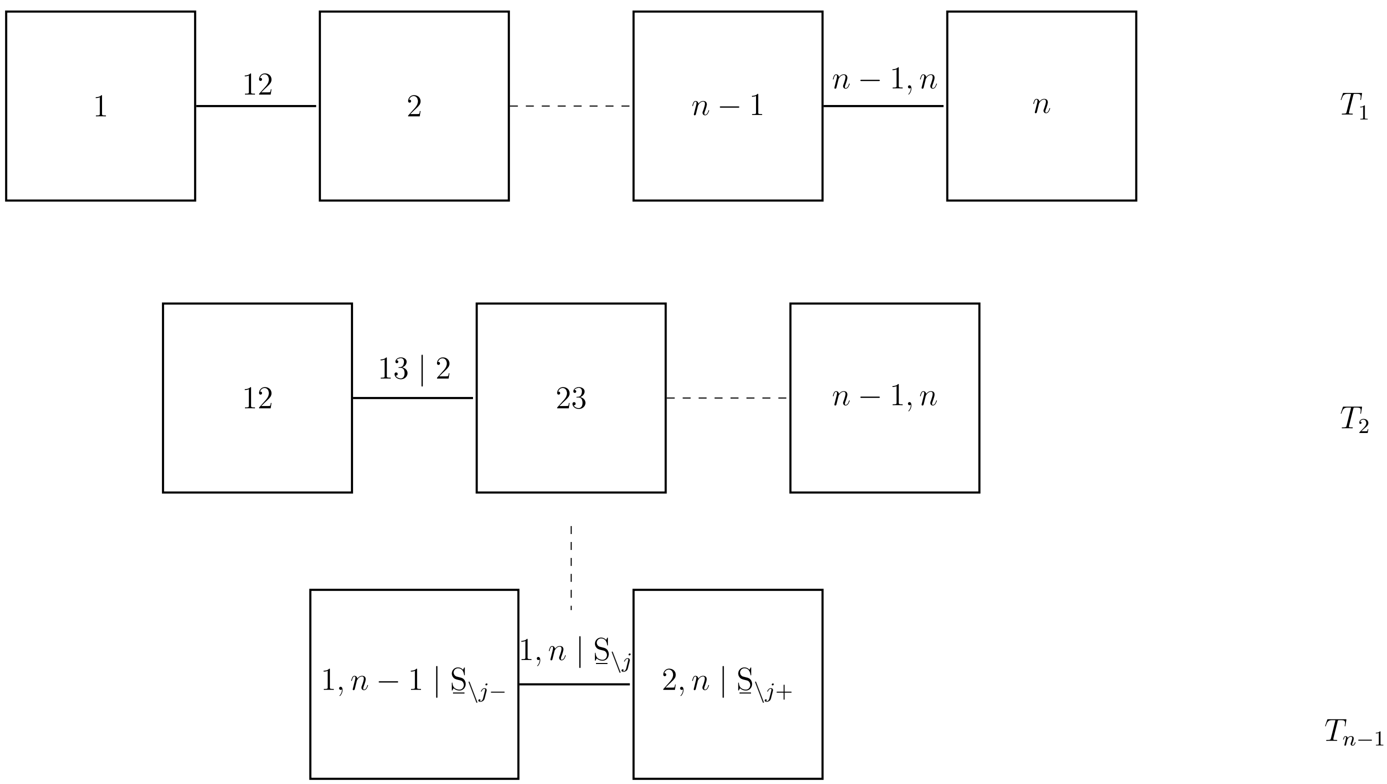
The D-vine PCC (which is depicted in figure 2.1) is constructed by $n-1$ trees, say $D = \{T_1, ..., T_{n-1}\}$, comprised of the edges $\xi(D) = \xi(T_1) \cup ... \cup \xi(T_{n-1})$, where $\xi(T_l)$ refers to the edges of the tree T_l. In the first tree T_1, the marginals $F(s_1), F(s_2), ..., F(s_n)$, are arranged as nodes in consecutive order, say $N(T_1) := \{1, 2, ..., n-1, n\}$, where the nodes are of degree two, meaning that no more than two edges is connected to each node. The corresponding edges join the adjacent nodes, such that $\xi(T_1) := \{12, 23, ..., (n-1, n)\}$. Next, the edges of the first tree $\xi(T_1)$ become the nodes of T_2, a process which is completed in a recursive manner, such that $N(T_{l+1}) = \xi(T_l)$, with the edges of each tree joining the adjacent nodes, and with the mutual elements between the nodes becoming the conditioning set. To express likelihood function (2.10) as a D-vine decomposition, we begin by calculating the marginals $F_1, ..., F_n$, where $F_t \sim Bernoulli(p_t)$, for $t = 1, ..., n$, with CDFs that are expressed as (2.9). Therefore, under assumptions (2.1) and (2.2), we have

$$F_t(s_t \mid X) = \begin{cases} 0, & \text{for} \quad s_t < 0 \\ 1 - P[\varepsilon_t \geq -\beta' x_{t-1} \mid X], & \text{for} \quad 0 \leq s_t < 1 \\ 1, & \text{for} \quad s_t \geq 1 \end{cases}, \quad t = 1, ..., n. \tag{2.14}$$

Once the marginals are obtained, the next step consists of evaluating the copulae in the first tree (i.e. $C_{12}(F_1, F_2), ..., C_{n-1,n}(F_{n-1}, F_n)$, corresponding to the edges $\xi(T_1)$). In the second tree, the copulae $C_{13|2}(F_{1|2}, F_{3|2}), ..., C_{n-2,n|n-1}(F_{n-2|n-1}, F_{n|n-1})$ are evaluated, next $C_{14|23}(F_{1|23}, F_{4|23}), ..., C_{n-3,n|n-2,n-1}(F_{n-3|n-2,n-1}, F_{n|n-2,n-1})$ in the third tree, and so on.

In the case of continuous variables, say $\{s^*(y_t) \in \mathbb{R}, t = 1, ..., n\}$, the construction of the D-vine involves an iterative copula evaluation process for the trees $T_1, ..., T_{n-1}$. This leads to $n(n-1)/2$

Figure 2.1: D-vine PCC for the n-variate case



Note: D-vine for a sample size n consists of $n-1$ trees. The first tree consists of the marginals ordered consecutively as nodes, with the edges connecting the adjacent nodes, and with the elements shared by the two nodes going in the conditioning set. The edges of each tree T_l become the nodes of the tree T_{l+1}. In this figure, $\underset{\sim}{S}_{\backslash i-}$ and $\underset{\sim}{S}_{\backslash j+}$ correspond to the elements $\underset{\sim}{S}_{\backslash j-} := \{2, 3, ..., n-2\}$ and $\underset{\sim}{S}_{\backslash j+} := \{3, ..., n-1\}$ respectively, with $\underset{\sim}{S}_{\backslash j} := \{2, 3, ..., n-1\}$.

copula evaluations, which corresponds to one copula evaluation for each edge [see Appendix]. On the other hand, for discrete variables, the conditional p.m.fs are expressed as in (2.12), which requires the evaluation of the following four copulae

$$\begin{aligned} &C^{++}_{t,j|\backslash\mathbf{j}}(F^{+}_{t|\backslash\mathbf{j}}, F^{+}_{j|\backslash\mathbf{j}}), \quad C^{+-}_{t,j|\backslash\mathbf{j}}(F^{+}_{t|\backslash\mathbf{j}}, F^{-}_{j|\backslash\mathbf{j}}), \\ &C^{-+}_{t,j|\backslash\mathbf{j}}(F^{-}_{t|\backslash\mathbf{j}}, F^{+}_{j|\backslash\mathbf{j}}), \quad C^{--}_{t,j|\backslash\mathbf{j}}(F^{-}_{t|\backslash\mathbf{j}}, F^{-}_{j|\backslash\mathbf{j}}), \end{aligned}$$

where $F^{+}_{t|\backslash\mathbf{j}} = P[s(y_t) \leq s_t \mid \underline{\mathrm{S}}^{\backslash j}_{t-1} = \underline{\mathrm{s}}^{\backslash j}_{t-1}, X]$ and $F^{-}_{t|\backslash\mathbf{j}} = P[s(y_t) \leq s_t - 1 \mid \underline{\mathrm{S}}^{\backslash j}_{t-1} = \underline{\mathrm{s}}^{\backslash j}_{t-1}, X]$. Henceforth, $4 \times n(n-1)/2$ bivariate copulae need to be evaluated in the case of discrete data. Let us express the joint p.m.f of the signs as follows

$$P_1[s(y_1) = s_1 \mid X] \times \prod_{t=2}^{n} P_{t|1:t-1}[s(y_t) = s_t \mid s(y_1) = s_1, ..., s(y_{t-1}) = s_{t-1}, X], \tag{2.15}$$

where following the result by Stoeber et al. (2013), if the D-vine is expressed as a vine-array $A = (\sigma_{lt})_{1 \leq l \leq t \leq n}$, where $l = 2, ..., n-1$ is the row with tree T_l, and column t has the permutation $\underline{\sigma}_{t-1} = (\sigma(1,t), ..., \sigma(t-1,t))$ of the previously added variables, such that

$$\begin{bmatrix} - & 12 & 23 & 34 & \cdots & n-1,n \\ & - & 13 \mid 2 & 24 \mid 3 & \cdots & n-2,n \mid n-1 \\ & & \ddots & \cdots & \cdots & \vdots \\ & & & - & 1,n-1 \mid \underline{\mathrm{S}}_{\backslash j-} & 2,n \mid \underline{\mathrm{S}}_{\backslash j+} \\ & & & & - & 1,n \mid \underline{\mathrm{S}}_{\backslash j} \\ & & & & & - \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 & 2 & 3 & \cdots & n-1 \\ & 2 & 1 & 2 & \cdots & n-2 \\ & & \ddots & \cdots & \cdots & \vdots \\ & & & n-2 & 1 & 2 \\ & & & & n-1 & 1 \\ & & & & & n \end{bmatrix}$$

then

$$P_{t|1:t-1}[s(y_t) = s_t \mid s(y_1) = s_1 : s(y_{t-1}) = s_{t-1}, X] = \left\{ \prod_{l=t-1}^{2} c_{\sigma_{lt}t,|\sigma_{1t},\ldots,\sigma_{t-1,t}} \right\} \times c_{\sigma_{lt}t} \times P_t[s(y_t) = s_t \mid X], \tag{2.16}$$

where following Joe (2014), the copula densities in expression (2.16) are calculated by

$$c_{t,j|\backslash\mathbf{j}} = \frac{C^{++}_{t,j|\backslash\mathbf{j}}(F^{+}_{t|\backslash\mathbf{j}}, F^{+}_{j|\backslash\mathbf{j}}) - C^{-+}_{t,j|\backslash\mathbf{j}}(F^{-}_{t|\backslash\mathbf{j}}, F^{+}_{j|\backslash\mathbf{j}}) - C^{+-}_{t,j|\backslash\mathbf{j}}(F^{+}_{t|\backslash\mathbf{j}}, F^{-}_{j|\backslash\mathbf{j}}) + C^{--}_{t,j|\backslash\mathbf{j}}(F^{-}_{t|\backslash\mathbf{j}}, F^{-}_{j|\backslash\mathbf{j}})}{P_{t|\backslash\mathbf{j}}[s(y_t) = s_t \mid \underline{\mathrm{S}}^{\backslash j}_{t-1} = \underline{\mathrm{s}}^{\backslash j}_{t-1}, X] P_{j|\backslash\mathbf{j}}[s(y_j) = s_j \mid \underline{\mathrm{S}}^{\backslash j}_{t-1} = \underline{\mathrm{s}}^{\backslash j}_{t-1}, X]}, \tag{2.17}$$

which results in the following proposition.

Proposition 3 *Let $A = (\sigma_{lt})_{1\leq l\leq t\leq n}$ be a D-vine array for the signs $s(y_1), ..., s(y_n)$. Under assumptions (2.1) and (2.2), let H_0 and H_1 be defined by (2.3) - (2.4),*

$$SL_n(\beta_1) = \sum_{t=2}^{n}\sum_{l=t-1}^{2} \ln c_{\sigma_{lt}t,|\sigma_{1t},...,\sigma_{t-1,t}} + \sum_{t=2}^{n} \ln c_{\sigma_{1t}t} + \sum_{t=1}^{n} s(y_t)a_t(\beta_1) > c_1(\beta_1),$$

where

$$a_t(\beta_1) = \ln\left\{\frac{1 - P_t[\varepsilon_t \leq -\beta' x_{t-1} \mid X]}{P_t[\varepsilon_t \leq -\beta' x_{t-1} \mid X]}\right\},$$

and suppose the constant $c_1(\beta_1)$ satisfies $P\left[SL_n(\beta_1) > c_1(\beta_1)\right] = \alpha$ under H_0, with $0 < \alpha < 1$. Then the test that rejects H_0 when

$$SL_n(\beta_1) > c_1(\beta_1) \tag{2.18}$$

is most powerful for testing H_0 against H_1 among level-α tests based on the signs $\big(s(y_1), ..., s(y_n)\big)'$.

Proof and algorithm: *See Appendix.*

As with the first chapter, under the null hypothesis the signs $s(y_1), ..., s(y_n)$ are i.i.d. according to Bernoulli $Bi(1, 0.5)$, with the distribution of $SL_n(\beta_1)$ only depending on the weights $a_t(\beta_1)$, without the presence of any nuisance parameters. Assumption (2.2) implies that tests based on $SL_n(\beta_1)$, such as the test given by (2.27), are distribution-free and robust against heteroskedasticity of unknown form. On the other hand, under the alternative hypothesis, the power function of the test depends on the form of the distribution of ε_t. A special case is where $\varepsilon_1, ..., \varepsilon_n$ are independently distributed according to $N(0, 1)$, which leads to the optimal test statistic assuming the following form

$$SL_n(\beta_1) = \sum_{t=2}^{n}\sum_{l=t-1}^{2} \ln c_{\sigma_{lt}t,|\sigma_{1t},...,\sigma_{t-1,t}} + \sum_{t=2}^{n} \ln c_{\sigma_{1t}t} + \sum_{t=1}^{n} s(y_t)a_t(\beta_1) > c_1(\beta_1),$$

where

$$a_t(\beta_1) = \ln\left\{\frac{\Phi(\beta' x_{t-1})}{1-\Phi(\beta' x_{t-1})}\right\},$$

where $\Phi(.)$ is the standard normal distribution function. The distibution of $SL_n(\beta_1)$ can be simulated under the null hypothesis with sufficient number of repliciations, and the critical values can be obatined to any degree of precision.

2.3.2 Testing general full coefficient hypothesis in non-linear regressions

We now consider a non-linear regression model

$$y_t = f(x_{t-1}, \beta) + \varepsilon_t,\ t = 1, ..., n, \tag{2.19}$$

where x_{t-1} is an observable $(k+1) \times 1$ vector of stochastic explanatory variables, such that $x_{t-1} = [1, x_{1,t-1}, ..., x_{k,t-1}]'$, $f(\cdot)$ is a scalar function, $\beta \in \mathbb{R}^{(k+1)}$ is an unknown vector of parameters and

$$\varepsilon_t \mid X \sim F_t(. \mid X)$$

where as before $F_t(. \mid X)$ is a distribution function and $X = [x_0, ..., x_{n-1}]$ is an $n \times (k+1)$ matrix. Once again, we suppose that the error terms process $\{\varepsilon_t, t = 1, 2, \cdots\}$ is a strict conditional mediangale, such that

$$P[\varepsilon_t > 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = P[\varepsilon_t < 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = \frac{1}{2}, \tag{2.20}$$

with

$$\boldsymbol{\varepsilon}_0 = \{\emptyset\}, \quad \boldsymbol{\varepsilon}_{t-1} = \{\varepsilon_1, \cdots, \varepsilon_{t-1}\}, \quad \text{for} \quad t \geq 2$$

and where (2.20) entails that $\varepsilon_t \mid X$ has no mass at zero, *i.e.* $P[\varepsilon_t = 0 \mid X]$=0 for all t. We do not require that the parameter vector β be identified.

We consider the problem of testing the null hypothesis

$$H(\beta_0) : \beta = \beta_0, \tag{2.21}$$

against the alternative hypothesis

$$H(\beta_1) : \beta = \beta_1, \tag{2.22}$$

We construct a test for $H(\beta_0)$ against $H(\beta_1)$ in a similar manner to Section 2.3.1, by first transforming model (2.19) to

$$\tilde{y}_t = g(x_{t-1}, \beta, \beta_0) + \varepsilon_t, \quad t = 1, ..., n$$

where $\tilde{y}_t = y_t - f(x_{t-1}, \beta_0)$ and $g(x_{t-1}, \beta, \beta_0) = f(x_{t-1}, \beta) - f(x_{t-1}, \beta_0)$. Notice that testing $H(\beta_0)$ against $H(\beta_1)$ is equivalent to testing

$$\bar{H}_0 : g(x_{t-1}, \beta, \beta_0) = 0, \quad \text{for } t = 1, ..., n$$

against the alternative

$$\bar{H}_A : g(x_{t-1}, \beta, \beta_0) = f(x_t, \beta_1) - f(x_t, \beta_0), \quad \text{for } t = 1, ..., n$$

For $\tilde{U}(n) = (s(\tilde{y}_1), ..., s(\tilde{y}_n))'$, where for $1 \leq t \leq n$

$$s(\tilde{y}_t) = \begin{cases} 1, \, if \;\, \tilde{y}_t \geq 0 \\ 0, \, if \;\, \tilde{y}_t < 0 \end{cases} .$$

As before, the joint p.m.f of the process of signs is expressed by

$$P_1[s(\tilde{y}_1) = \tilde{s}_1 \mid X] \times \prod_{t=2}^{n} P_{t|1:t-1}[s(\tilde{y}_t) = \tilde{s}_t \mid s(\tilde{y}_1) = \tilde{s}_1, ..., s(\tilde{y}_{t-1}) = \tilde{s}_{t-1}, X]. \tag{2.23}$$

Furthermore, the D-vine-array $\tilde{A} = (\tilde{\sigma}_{lt})_{1 \leq l \leq t \leq n}$, is such that $l = 2, ..., n-1$ is the row with tree T_l, and column t has the permutation $\underline{\tilde{\sigma}}_{t-1} = (\tilde{\sigma}_{1t}, ..., \tilde{\sigma}_{t-1,t})$ of the previously added variables. Then

$$P_{t|1:t-1}[s(\tilde{y}_t) = \tilde{s}_t \mid s(\tilde{y}_1) = \tilde{s}_1, ..., s(\tilde{y}_{t-1}) = \tilde{s}_{t-1}, X] = \left\{ \prod_{l=t-1}^{2} c_{\tilde{\sigma}_{lt}t,|\tilde{\sigma}_{1t},...,\tilde{\sigma}_{t-1,t}} \right\} \times c_{\tilde{\sigma}_{lt}t} \times P_t[s(\tilde{y}_t) = \tilde{s}_t \mid X], \tag{2.24}$$

which leads to the following corollary.

Corollary 3 *Let $\tilde{A} = (\tilde{\sigma}_{lt})_{1 \leq l \leq t \leq n}$ be a D-vine array for the signs $s(\tilde{y}_1), ..., s(\tilde{y}_n)$. Under assumptions (2.19) and (2.2), let $H(\beta_0)$ and $H(\beta_1)$ be defined by (2.21) - (2.22),*

$$SN_n(\beta_0 \mid \beta_1) = \sum_{t=2}^{n} \sum_{l=t-1}^{2} \ln c_{\tilde{\sigma}_{lt}t,|\tilde{\sigma}_{1t},...,\tilde{\sigma}_{t-1,t}} + \sum_{t=2}^{n} \ln c_{\tilde{\sigma}_{1t}t} + \sum_{t=1}^{n} s(y_t - f(x_{t-1}, \beta_0))\tilde{a}_t(\beta_0 \mid \beta_1) > c_1(\beta_0, \beta_1),$$

where

$$\tilde{a}_t(\beta_0 \mid \beta_1) = \ln \left\{ \frac{1 - p_t[x_{t-1}, \beta_0, \beta_1 \mid X]}{p_t[x_{t-1}, \beta_0, \beta_1 \mid X]} \right\}, \quad p_t[x_{t-1}, \beta_0, \beta_1 \mid X] = P_t[\varepsilon_t \leq f(x_{t-1}, \beta_0) - f(x_{t-1}, \beta_1) \mid X]$$

and suppose the constant $c_1(\beta_0, \beta_1)$ satisfies the constraint $P\left[SN_n(\beta_0 \mid \beta_1) > c_1(\beta_0, \beta_1)\right] = \alpha$ under $H(\beta_0)$, with $0 < \alpha < 1$. Then the test that rejects $H(\beta_0)$ when

$$SN_n(\beta_0 \mid \beta_1) > c_1(\beta_0, \beta_1) \tag{2.25}$$

is most powerful for testing $H(\beta_0)$ against $H(\beta_1)$ among level-α tests based on the signs $\big(s(\tilde{y}_1), ..., s(\tilde{y}_n)\big)'$.

Consider a linear function $f(x_{t-1}, \beta) = \beta' x_{t-1}$, and assume that under the alternative hypothesis ε_t for $t = 1, ..., n$ follows a standard normal distribution (i.e. $\varepsilon_t \sim N(0, 1)$). Then the statistic for testing $H(\beta_0)$ against the alternative $H(\beta_1)$ is given by

$$SN_n(\beta_0 \mid \beta_1) = \sum_{t=2}^{n} \sum_{l=t-1}^{2} \ln c_{\tilde{\delta}_{lt}t,|\tilde{\delta}_{1t},...,\tilde{\delta}_{t-1,t}} + \sum_{t=2}^{n} \ln c_{\tilde{\delta}_{1t}t} + \sum_{t=1}^{n} s(y_t - \beta_0' x_{t-1})\tilde{\delta}_t(\beta_0 \mid \beta_1) > c_1(\beta_0, \beta_1),$$

where

$$\tilde{a}_t(\beta_0 \mid \beta_1) = \ln \left\{ \frac{\Phi((\beta_1 - \beta_0)' x_{t-1})}{1 - \Phi((\beta_1 - \beta_0)' x_{t-1})} \right\},$$

such that $\Phi(.)$ is the standard normal distribution function. As in Section 2.3, the distribution of $SN_n(\beta_0 \mid \beta_1)$ can be simulated under the null hypothesis with sufficient number of replications and the relevant critical values can be obtained to any degree of precision.

2.4 Estimation

In this Section, we first consider the issue of estimating the bivariate copulae in the D-vine decomposition and suggest a sequential estimation strategy for the parameters of the copulae. We then turn our attention to the problem of selecting a class of parametric bivariate copulae. The choice of the latter has an important implication on introducing dependence to the vector of signs.

2.4.1 Sequential estimation of the D-vine

The calculation of the test statistics in Section 2.3 requires four bivariate copula evaluations at $n(n-1)/2$ distinct points, leading to a total of $2n(n-1)$ copula evaluations. The estimation of the D-vine is often facilitated with the maximum likelihood (MLE). However, as the latter requires optimization with respect to at least $2n(n-1)$ copula parameters, sequential estimation procedures are favored for faster computation times, with the caveat that the increased speed comes at the cost of efficiency. Furthermore, the sequential estimates may be provided as starting points for the simultaneous numerical optimization using MLE [see Czado et al. (2012); Haff (2012); Dissmann et al. (2013)]. We assume that the copulae are specified parametrically, given by an appropriate parameter (vector). More specifically, let $\boldsymbol{\theta}_l = (\theta'_{1,k}, ..., \theta'_{n-l,k})'$ be the set of all the parameters to be estimated for tree l, $l = 1, ..., n-1$ of the D-vine, with $k = l-1$ conditioning variables. Therefore, $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, ..., \boldsymbol{\theta}'_{n-1})'$ is the entire set of the parameters that needs to be estimated for the D-vine decomposition. To estimate the parameter vector $\boldsymbol{\theta}$, we follow a sequential estimation strategy proposed by Czado et al. (2012), whereby first, the parameters of the unconditional bivariate copulae are estimated. These parameters are then utilized as means of estimating the parameters of bivariate copulae with a single conditioning variable. The latter are then used to estimate the pair-copulae with two conditioning variables, and so on. This bivariate copula estimation approach is continued sequentially until all parameters have been estimated.

In the first step, the marginals are obtained by computing the Bernoulli CDFs (2.14) using an arbitrary distribution, such as the standard normal distribution considered in Section 2.3. The second step of the process involves estimating the parameters of the unconditional copula, by fixing the marginals with their aforementioned estimates and maximizing the bivariate likelihood corresponding to each copula in each tree l to obtain $\hat{\boldsymbol{\theta}}_l = (\hat{\theta}'_{1,k}, ..., \hat{\theta}'_{n-l,k})$ for $l = 1, ..., n-1$ and $k = l - 1$. As all the variables are discrete, the log-likelihood function, say, for the unconditional copula $C_{t,t+1}$ for $t = 1, ..., n-1$ for the signs $(s(y_{i,t}), s(y_{i,t+1}))$, $i = 1, ..., n-1$ is expressed as

$$L(\theta_{t,0}) = \sum_{i=1}^{n-1} \log \left\{ \sum_{\{a_1,a_2\} \in \{-,+\}^2} (-1)^{a_j} C_{t,t+1} \left(F_t(s_{i,t}^{a_1} \mid X; \hat{\beta}_{\mathbf{1}}), F_{t+1}(s_{i,t+1}^{a_2} \mid X; \hat{\beta}_{\mathbf{1}}); \theta_{t,0} \right) \right\}.$$

The estimate of the copula parameter, $\hat{\theta}_{t,0}$ for $t = 1, ..., n-1$, is then obtained as follows

$$\hat{\theta}_{t,0} = \arg\max_{\theta_{t,0}} L(\theta_{t,0}),$$

which under regularity conditions solves

$$\frac{\partial L(\theta_{t,0})}{\partial \theta_{t,0}} = 0.$$

Let us illustrate this process with an example: once the marginals have been obtained, the next step involves estimating the parameters $\theta_{t,0}$ for $t = 1, ..., n-1$ of the unconditional copulae. Next, we are interested in estimating $\theta_{t,1}$ for $t = 1, ..., n-2$. Define

$$\hat{u}_{t|t+1} = F_{t|t+1} \left(s_t \mid s_{t+1}, X; \hat{\theta}_{t,0} \right),$$

and

$$\hat{v}_{t+2|t+1} = F_{t+2|t+1} \left(s_{t+2} \mid s_{t+1}, X; \hat{\theta}_{t+1,0} \right),$$

for $t = 1, ..., n-2$. The data $\hat{u}_{t|t+1}$ and $\hat{v}_{t+2|t+1}$ is then used to estimate the parameters $\theta_{t,1}$ for $t = 1, ..., n-2$, denoted by $\hat{\theta}_{t,1}$. This procedure is repeated sequentially until all parameters have been estimated. Haff et al. (2010) show that under regularity conditions, the sequential estimates are asymptotically normal; however, as noted earlier their asymptotic covariance is "intractable"and the faster computation time comes at the cost of efficiency. Therefore, the sequential estimates can be utilized as the starting values of the high-dimensional MLE.

It is worth mentioning that although estimating the parameters for each of the $2n(n-1)$ bivariate copulae does not present any significant issues in finite samples, this approach is not feasible as $n \to \infty$, since the estimation procedure, especially using the MLE approach becomes computationally burdensome in large samples. However, it is immediately evident that when the process y_t is stationary, the parameters of the copulae for each tree l are invariant to time shifts and the problem of obtaining $\hat{\boldsymbol{\theta}}_l = (\hat{\theta}'_{1,k}, ..., \hat{\theta}'_{n-l,k})$ for $l = 1, \cdots, n-1$ and $k = l-1$, reduces to computing a single parameter vector $\hat{\theta}'_k$. In other words, we would only have to estimate $\hat{\boldsymbol{\theta}}_l = \hat{\theta}'_k$ for $l = 1, ..., n-1$ and $k = l-1$, which reduces the problem to $n-1$ parameter vector estimations. This number can further be reduced to one parameter, if we consider a 1-truncated D-vine [see Section 2.4.3].

Another approach for estimating one-parameter pair-copulae in the sequential estimation procedure for copula families with a known relationship to Kendall's τ consists of inverting the empirical Kendall's τ based on, say, $\hat{u}_t$ and $\hat{u}_{t+1}$ for $t = 1, ..., n-1$ for the edges of the first tree. However, we provide a caveat that the Kendall's τ of discrete data does not correspond to the Kendall's τ of the bivariate copulae [see Denuit and Lambert (2005)]. Denuit and Lambert (2005) show that by continuous extension of the discrete variables with a perturbation with values in $[0, 1]$, the continuous features of Kendall's τ are adaptable to discrete data. In other words

$$\tau(s^*(y_t), s^*(y_{t+1})) = 4 \int\int_{[0,1]^2} C^*_{t,t+1}\left(\hat{u}_t, \hat{u}_{t+1}\right) dC^*_{t,t+1}\left(\hat{u}_t, \hat{u}_{t+1}\right) - 1$$

for $t = 1, ..., n-1$, such that $\hat{u}_t = F\left(s^*_t; \hat{\beta}_1\right)$,

$$s^*(y_t) = s(y_t) + U - 1$$

where U is a continuous random variable in $[0, 1]$. A natural choice for U is the uniform distribution.

2.4.2 Selection of the copula family

Many different classes of parametric bivariate copulae have extensively been studied and reviewed by Joe (2014) and others that can fit within the framework of the PCC POS-based tests. These include the Archimedean, elliptical, extreme value or max-id copula families that can specify the dependence structure of the vector of signs. As the dependence is introduced by the copula family, the type and the degree of the dependence between the signs depends on the choice of the copula. The literature surrounding the goodness-of-fit of copulae is extensive and has been analyzed by Genest et al. (2006), Genest et al. (2009), and Berg (2009), among many others. Genest et al. (2009) categorize goodness-of-fit tests into three broad categories: procedures for testing particular dependence structures such as Gaussian or Clayton family; procedures that may be used for any classes of copulae, but require a strategic choice for their implementation; and finally, the so-called "blanket tests"that apply to all classes of copulae and require no strategic choice for their use. A simple procedure proposed by Joe (1997) involves specifying the Akaike information critetion (AIC) to different copulae and using it as a copula selection criterion, which is particularly attractive as it allows for the automation of the copula selection process [see. Czado et al. (2012)]. The AIC specified to the copulae of, say, the first tree of the D-vine, can be expressed as follows

$$AIC = -2\sum_{i=1}^{t} \log c_{t,t+1}(\hat{u}_{i,t}, \hat{u}_{i,t+1}; \hat{\theta}_{1,k}) + 2l$$

for $t = 1, ..., n-1$ and $k = 1, ..., n-1$, and where l is the number of parameters $\theta_{1,k}$. Panagiotelis et al. (2012) suggest that while dependence structures such as tail dependence are weak in discrete data, the choice of the copulae could still have a significant effect on the joint pmf of the signs. They considered Gaussian, Clayton and Gumbel copulae in constructing the D-vines for Bernoulli margins by keeping the marginal probabilities and dependence constant, and have found that in the case where the probabilities of zero marginals and joint probabilities in the data is high, preference goes to the use of the Gumbel copula over the other two alternatives.

In the earlier Section, we had considered standard normal distribution as the marginals. We further choose the Gaussian copula for the computation of the PCC-POS-based tests, due to its desirable properties, such as probabilistic interpretability, flexibility, and a wide range of dependence [see

Durante and Sempi (2010)]. Our simulation results show that the use of the Gaussian copula for the PCC-POS-based tests under different distributions and heteroskedasticities yields power that is superior to the other tests considered in our study in most circumstances. The normal copula model is equivalent to the bivariate normal distributions, as

$$\begin{aligned} F(s_t, s_{t+1}) &= C_\rho^\Phi(\Phi(s_t), \Phi(s_{t+1})) \\ &= \Phi_2\left(\Phi^{-1}(\Phi(s_t)), \Phi^{-1}(\Phi(s_{t+1})); \rho_{t,t+1}\right) \\ &= \Phi(s_t, s_{t+1}), \end{aligned}$$

for the first tree, and where $\rho_{t,t+1}$ is the correlation coefficient of the bivariate standard normal distribution. In other words

$$\begin{aligned} C_{t,t+1}^\Phi\left(\Phi(s_t), \Phi(s_{t+1}); \rho_{t,t+1}\right) &= \frac{1}{2\pi\sqrt{(1-\rho_{t,t+1}^2)}} \times \\ & \int_{-\infty}^{\Phi^{-1}(\Phi(s_t))} \int_{-\infty}^{\Phi^{-1}(\Phi(s_{t+1}))} e^{\left\{-\frac{\Phi(s_t)^2 - 2\rho_{t,t+1}\Phi(s_t)\Phi(s_{t+1}) + \Phi(s_{t+1})^2}{2(1-\rho_{t,t+1}^2)}\right\}} d\Phi(s_t) d\Phi(s_{t+1}). \end{aligned}$$

2.4.3 Truncated D-vines

Following Joe (2014), we refer to a D-vine as a p-truncated D-vine, if the copulae in the trees $T_{p+1}, ..., T_n$ are $C^\perp$, where by definition

$$C^\perp(u_1, ..., u_n) = \prod_{t=1}^{n} u_t, \quad \text{with} \quad (U_1, ..., U_n) \sim U(0,1),$$

implying $U_1 \perp U_2 \perp ... \perp U_n$. The POS-based tests in Chapter 1 can be regarded as a special case of the PCC-POS based tests, whereby the former can be constructed by a 1-truncated D-vine, which only depends on $C_{12}, C_{23}, ..., C_{n-1,n}$, or rather $C_{12}, C_{\sigma_{13}3}, ..., , C_{\sigma_{1n}n}$ in a vine array representation, given the following vine array

$$\begin{bmatrix} - & 12 & 23 & 34 & \cdots & n-1,n \\ & - & 13\mid 2^{\perp} & 24\mid 3^{\perp} & \cdots & n-2,n\mid n-1^{\perp} \\ & & \ddots & \cdots & \cdots & \vdots \\ & & & - & 1,n-1\mid \underline{\mathrm{S}}^{\perp}_{\backslash j-} & 2,n\mid \underline{\mathrm{S}}^{\perp}_{\backslash j+} \\ & & & & - & 1,n\mid \underline{\mathrm{S}}^{\perp}_{\backslash j} \\ & & & & & - \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 & 2 & 3 & \cdots & n-1 \\ & 2 & 1^{\perp} & 2^{\perp} & \cdots & n-2^{\perp} \\ & & \ddots & \cdots & \cdots & \vdots \\ & & & n-2 & 1^{\perp} & 2^{\perp} \\ & & & & n-1 & 1^{\perp} \\ & & & & & n \end{bmatrix}$$

Similarly, a 2-truncated D-vine, depends on the copulae $C_{12}, C_{23}, ..., C_{n-1,n}$ and $C_{13|2}, C_{24|3}, ..., C_{n-2,n|n-1}$ or $C_{\sigma_{1t}}t$ for $t = 2, ..., n$ and $C_{\sigma_{2t}t|\sigma_{1t}}$ for $t = 3, ..., n$ using the vine array representation. Therefore, for a p-truncated D-vine, (2.24) is modified to

$$\begin{aligned} P_{t|1:t-1}[s(\tilde{y}_t) = \tilde{s}_t \mid s(\tilde{y}_1) = \tilde{s}_1, ..., s(\tilde{y}_{t-1}) = \tilde{s}_{t-1}, X] &= P_{t|1:p}[s(\tilde{y}_t) = \tilde{s}_t \mid s(\tilde{y}_1) = \tilde{s}_1, ..., s(\tilde{y}_p) = \tilde{s}_p, X] \\ &= \left\{ \prod_{l=p\wedge(t-1)}^{2} c_{\tilde{\sigma}_{lt}t,|\tilde{\sigma}_{1t},...,\tilde{\sigma}_{t-1,t}} \right\} \times c_{\tilde{\sigma}_{lt}t} \times P_t[s(\tilde{y}_t) = \tilde{s}_t \mid X], \end{aligned} \tag{2.26}$$

for $t - 1 \geq p$.

Corollary 4 *Let $\tilde{A} = (\tilde{\sigma}_{lt})_{1\leq l\leq t\leq n}$ be a D-vine array for the signs $s(\tilde{y}_1), ..., s(\tilde{y}_n)$, where the signs $\{s(\tilde{y}_t)\}_{t=0}^{\infty}$ follow a Markov process of order p. Under assumptions (2.19) and (2.2), let $H(\beta_0)$ and $H(\beta_1)$ be defined by (2.21) - (2.22),*

$$SN_n(\beta_0 \mid \beta_1) = \sum_{t=2}^{n} \sum_{l=p\wedge(t-1)}^{2} \ln c_{\tilde{\sigma}_{lt}t,|\tilde{\sigma}_{1t},...,\tilde{\sigma}_{t-1,t}} + \sum_{t=2}^{n} \ln c_{\tilde{\sigma}_{1t}t} + \sum_{t=1}^{n} s(y_t - f(x_{t-1}, \beta_0))\tilde{a}_t(\beta_0 \mid \beta_1) > c_1(\beta_0, \beta_1),$$

where

$$\tilde{a}_t(\beta_0 \mid \beta_1) = \ln\left\{\frac{1 - p_t[x_{t-1}, \beta_0, \beta_1 \mid X]}{p_t[x_{t-1}, \beta_0, \beta_1 \mid X]}\right\}, \quad p_t[x_{t-1}, \beta_0, \beta_1 \mid X] = P_t[\varepsilon_t \leq f(x_{t-1}, \beta_0) - f(x_{t-1}, \beta_1) \mid X]$$

and suppose the constant $c_1(\beta_0, \beta_1)$ *satisfies the constraint* $P\left[SN_n(\beta_0 \mid \beta_1) > c_1(\beta_0, \beta_1)\right] = \alpha$ *under* $H(\beta_0)$, *with* $0 < \alpha < 1$. *Then the test that rejects* $H(\beta_0)$ *when*

$$SN_n(\beta_0 \mid \beta_1) > c_1(\beta_0, \beta_1) \tag{2.27}$$

is most powerful for testing $H(\beta_0)$ *against* $H(\beta_1)$ *among level-*α *tests based on the signs* $\big(s(\tilde{y}_1), ..., s(\tilde{y}_n)\big)'$.

2.5 Choice of the optimal alternative hypothesis

In this Section, we follow Dufour and Taamouti (2010a) by first showing the analytical derivation of the power envelope function of the PCC-POS-based tests. We then suggest using simulations as means of approximating the said function, by showing the difficulty of inverting the latter to find the optimal alternative. Thereafter, we propose an adaptive approach based on the split-sample technique to choose an alternative which has a power function close to that of the power envelope.

2.5.1 Power envelope of PCC-POS tests

We make the assertion that point-optimal tests trace out the power envelope (i.e. the maximum attainable power) for any given testing problem [see King (1987)]. However, in practice the alternative hypothesis β_1 is unknown and a problem consists of finding an approximation for it, such that the power function is maximized and is close to that of the power envelope. Following Dufour and Taamouti (2010a) and Dufour and Jasiak (2001), we propose an adaptive approach based on the split-sample technique to choose an alternative β_1 that yields the greatest power function. The details of the split-sample technique can be found in Section 2.5.2. We follow Dufour and Taamouti (2010a) by showing the analytical derivation of the power envelope of the PCC-POS tests for predictive regressions, which can be purposed as a benchmark for comparing the power functions of the PCC-POS tests for different sample splits.

We have shown in Section 2.3.2 that the PCC-POS test is a function of β_1. In other words,

$$SN_n(\beta_0 \mid \beta_1) = \sum_{t=2}^{n}\sum_{l=t-1}^{2} \ln c_{\tilde{\sigma}_{lt}t|\tilde{\sigma}_{1t},\ldots,\tilde{\sigma}_{t-1,t}} + \sum_{t=2}^{n} \ln c_{\tilde{\sigma}_{1t}t} + \sum_{t=1}^{n} \ln\left\{\frac{1-p_t[x_{t-1},\beta_0,\beta_1 \mid X]}{p_t[x_{t-1},\beta_0,\beta_1 \mid X]}\right\} s(y_t - f(x_{t-1},\beta_0)).$$

which in turn implies that its power function, say $\Pi(\beta_0, \beta_1)$, is also a function of β_1

$$\Pi(\beta_0,\beta_1) = P[SN_n(\beta_0 \mid \beta_1) > c_1(\beta_0,\beta_1) \mid H(\beta_1)]$$

where $c_1(\beta_0, \beta_1)$ is the smallest constant that satisfies $P[SN_n(\beta_0 \mid \beta_1) > c_1(\beta_0,\beta_1) \mid H(\beta_0)] \leq \alpha$, and where α is an arbitrary significance level. Theorem 2 provides the theoretical results for the power function of the PCC-POS tests.

Theorem 2 *Under assumption (2.2) and and given model (2.19), and further under the condition that $s(\tilde{y}_1), \ldots, s(\tilde{y}_t)$ follow a Regularity Markov Type process, the power function of $SN_n(\beta_0 \mid \beta_1)$ is given by*

$$\Pi(\beta_0,\beta_1) = P[SN_n(\beta_0 \mid \beta_1) > c_1(\beta_0,\beta_1) \mid X] = \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} \frac{\mathrm{Im}\{\exp(iuc_1(\beta_0,\beta_1))\phi_{SN_n}(u)\}}{u} du$$

$\forall u \in \mathbb{R}$, $i = \sqrt{-1}$, and with $\mathrm{Im}\{z\}$ denoting the imaginary part of the complex number z. $\phi_{SN_n}(u)$ is given by

$$\phi_{SN_n}(u) = \prod_{t=1}^{n}\left(\mathbb{E}_X\left[\exp\left(iu\left\{R_{t,t-1} + \ln\left\{\frac{1-p_t[x_{t-1},\beta_0,\beta_1 \mid X]}{p_t[x_{t-1},\beta_0,\beta_1 \mid X]}\right\} s(\tilde{y}_t)\right\}\right)\right] + \rho_t(u)\right),$$

where $R_{1,0} = 0$, $R_{t,t-1} = \sum_{l=t-1}^{2} \ln c_{\tilde{\sigma}_{lt}t|\tilde{\sigma}_{1t},\ldots,\tilde{\sigma}_{t-1,t}} + \ln c_{\tilde{\sigma}_{1t}t}$ for $t = 2, \ldots, n$, such that for D-vine-array $\tilde{A} = (\tilde{\sigma}_{lt})_{1\leq l\leq t\leq n}$, $l = 2, \ldots, n-1$ is the row with tree T_l, and column t has the permutation $\underline{\tilde{\sigma}}_{t-1} = (\tilde{\sigma}_{1t}, \ldots, \tilde{\sigma}_{t-1,t})$ of the previously added variables, $p_t[x_{t-1},\beta_0,\beta_1 \mid X] = P_t[\varepsilon_t \leq f(x_{t-1},\beta_0) - f(x_{t-1},\beta_1) \mid X]$, $\underline{\tilde{S}}_{t-1} = s(y_{t-1} - f(x_{t-2},\beta_0)), \ldots, s(y_1 - f(x_0,\beta_0))$ and

$$p_t[x_{t-1},\beta_0,\beta_1 \mid \underline{\tilde{S}}_{t-1} = \underline{\tilde{s}}_{t-1}, X] = \left\{\prod_{l=t-1}^{2} c_{\tilde{\sigma}_{lt}t,|\tilde{\sigma}_{1t},\ldots,\tilde{\sigma}_{t-1,t}}\right\} \times c_{\tilde{\sigma}_{lt}t} \times P_t[s(y_t - f(x_{t-1},\beta_0)) = j \mid X]$$

Finally, $c_1(\beta_0, \beta_1)$ *is the smallest constant that satisfies* $P[SN_n(\beta_0 \mid \beta_1) > c_1(\beta_0, \beta_1) \mid H(\beta_0)] \leq \alpha$, *where* α *is an arbitrary significance level.*

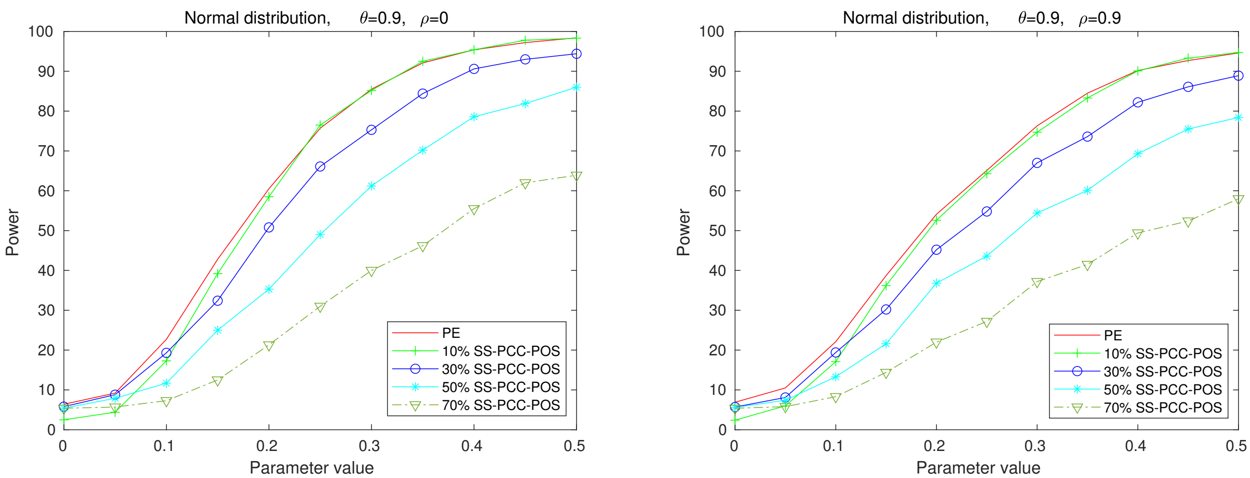
Proof: *See Appendix.*

Under the assumption that the signs follow an RMT-process, $\rho_t(u)$ can be estimated using the results from Theorem 2 of Heinrich (1982). Given that point-optimal tests are optimal at a specific point in the alternative parameter space, the power envelope of the PCC-POS tests, say $\bar{\Pi}(\beta_1)$, is obtained for values of β, such that $\{\beta : \beta = \beta_1, \forall \beta_1 \in \mathbb{R}^{(k+1)}\}$. Finding values of β_1 for a PCC-POS test at level α, with a power function that is close to the power envelope can be achieved by inverting the power envelope function. However, in a much simpler case of POS tests for i.n.i.d data, Dufour and Taamouti (2010a) show that the inversion of the power function is not a straightforward task and obtaining an exact solution is not feasible. Therefore, simulations are used as means of approximating the power envelope function and finding the optimal alternative for the PCC-POS test.

2.5.2 Split-sample technique for choosing the optimal alternative

As we have noted in the earlier Section, the power function of the PCC-POS test statistic depends on the alternative β_1, which in practice is unknown and needs to be approximated. To make size control easier and to choose an approximation to β_1 such that the power function of the test statistic is close to that of the power envelope, we follow Dufour and Taamouti (2010a) by proposing an adaptive approach based on the split-sample technique for choosing the alternative. For an extensive review of adaptive statistical methods, we refer the reader to O'Gorman (2004). Furthermore, the application of the split-sample technique in parametric settings can be studied by consulting Dufour and Taamouti (2003) and Dufour et al. (2008). The split-sample technique involves splitting a sample of size n into two independent subsamples, say n_1 and n_2, such that $n = n_1 + n_2$. The first subsample is then used to estimate the alternative β_1, while the other is purposed for computing the PCC-POS test statistic. Assuming that $f(x_{t-1}, \beta) = x'_{t-1}\beta$, the

Figure 2.2: Power comparisons: different split-samples. Normal error distributions with different values of ρ in (2.33) and $\theta = 0.9$ in (2.32)



Note: These figures compare the power envelope the PCC-POS test statistic using different split-samples: 10%, 30%, 50%, 70%. "PE"refers to the power envelope of the PCC-POS test.

alternative β_1 can be estimated using OLS

$$\hat{\beta}_{(1)} = (X'_{(1)} X_{(1)})^{-1} X'_{(1)} y_{(1)}.$$

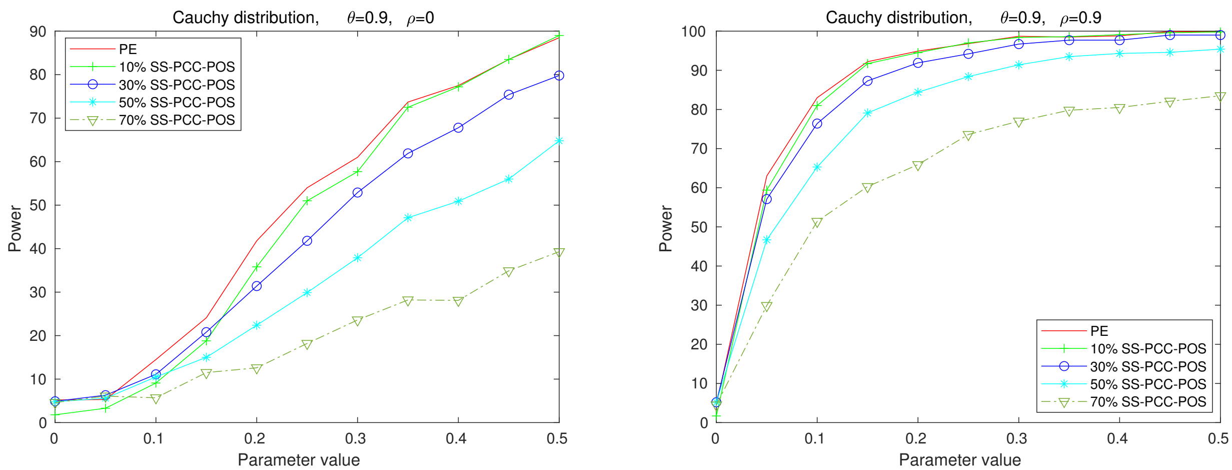
We provide a caveat that the OLS estimator is sensitive to extreme outliers, which motivates the use of robust estimators [see. Maronna et al. (2019) for a review of robust estimators]. Using $\hat{\beta}_{(1)}$ and the observations in the second independent subsample, we compute the test-statistic as follows

$$\begin{aligned} SN_n(\beta_0 \mid \beta_{(1)}) \;\; &= \;\; \sum_{t=n_1+2}^{n} \sum_{l=t-1}^{2} \ln c_{\tilde{\sigma}_{lt}t|\tilde{\sigma}_{(n_1+1)t},\dots,\tilde{\sigma}_{t-1,t}} + \sum_{t=n_1+2}^{n} \ln c_{\tilde{\sigma}_{(n_1+1)t}t} + \\ &\quad \sum_{t=n_1+1}^{n} \ln \left\{ \frac{1 - p_t[x_{t-1}, \beta_0, \beta_{(1)} \mid X]}{p_t[x_{t-1}, \beta_0, \beta_{(1)} \mid X]} \right\} s(y_t - x'_{t-1}\beta_0). \end{aligned}$$

where for $t = n_1+2, \dots, n$ and D-vine-array $\tilde{A} = (\tilde{\sigma}_{lt})_{1 \le l \le t \le n}$, $l = n_1+1, \dots, n-1$ is the row with tree T_l, and column t has the permutation $\underset{\sim}{\tilde{\sigma}}_{t-1} = (\tilde{\sigma}_{(n_1+1)t}, \dots, \tilde{\sigma}_{t-1,t})$ of the previously added variables, $p_t[x_{t-1}, \beta_0, \beta_{(1)} \mid X] = P_t[\varepsilon_t \le x'_{t-1}(\beta_0 - \beta_1) \mid X]$, and $\underset{\sim}{\tilde{S}}_{t-1} = s(y_{t-1} - x'_{t-2}\beta_0), \dots, s(y_{n_1+2} - x'_{n_1+1}\beta_0)$.

The choices for the subsamples n_1 and n_2 can be arbitrary. However, our simulations show that

Figure 2.3: Power comparisons: different split-samples. Cauchy error distributions with different values of ρ in (2.33) and $\theta = 0.9$ in (2.32)



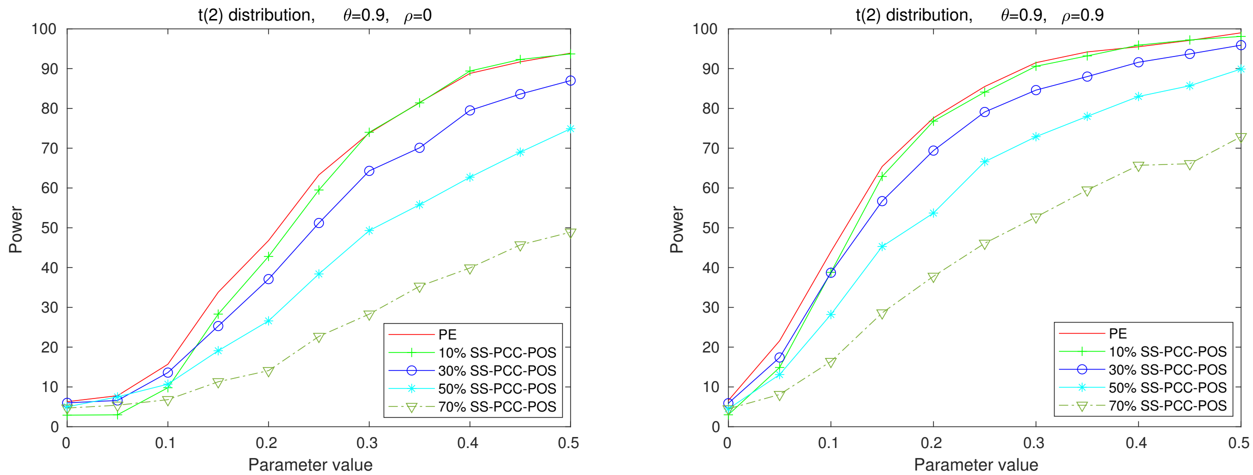
Note: These figures compare the power envelope the PCC-POS test statistic using different split-samples: 10%, 30%, 50%, 70%. "PE"refers to the power envelope of the PCC-POS test.

the proportions of observations retained for estimating the alternative and computing the PCC-POS test statistic has an impact on the power of the test. We find that the power function of the split-sample PCC-POS test (SS-PCC-POS test hereafter) is closest to that of the power envelope, when a relatively small number of observations is retained for estimating the alternative, with the rest used for computing the test statistic . These findings are in line with Dufour and Taamouti (2010a). Specifically, by considering all the DGPs in our simulations study, we find that the subsamples n_1 and n_2 must in turn contain roughly 10% and 90% of the observations of the entire sample respectively.

2.6 PCC-POS confidence regions

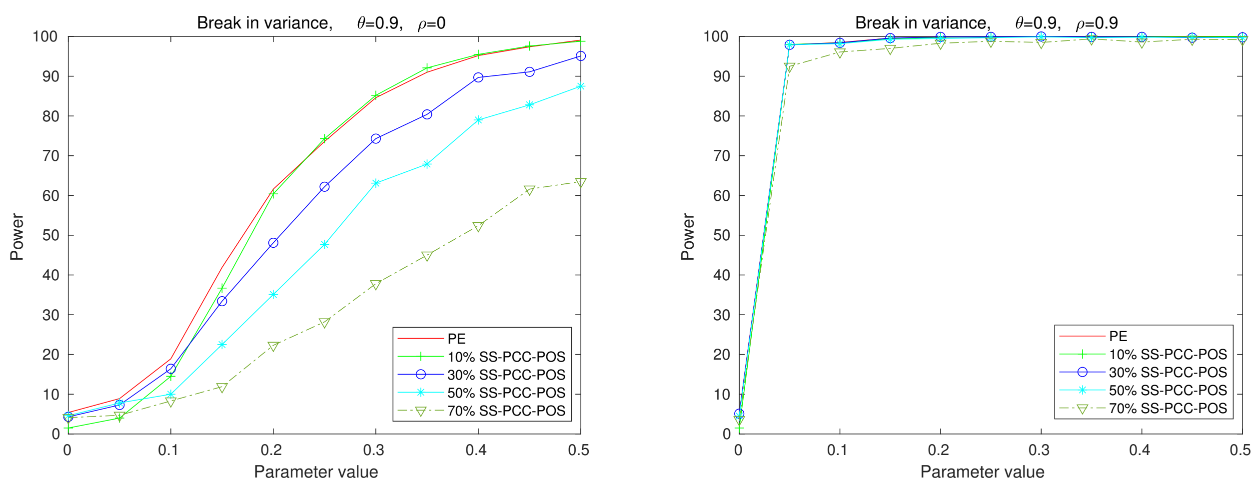
In this Section, we lay out the theoretical framework for building confidence regions for a vector (sub-vector) of the unknown parameters β, say $C_\beta(\alpha)$, at a given significance level α, using the proposed PCC-POS tests. Consider model (2.19) such that $f(x_{t-1}, \beta) = x'_{t-1}\beta$. Suppose we wish

Figure 2.4: Power comparisons: different split-samples. Student's t error distributions with 2 degrees of freedom [i.e $t(2)$] with different values of ρ in (2.33) and $\theta = 0.9$ in (2.32)



Note: These figures compare the power envelope the PCC-POS test statistic using different split-samples: 10%, 30%, 50%, 70%. "PE"refers to the power envelope of the PCC-POS test.

Figure 2.5: Power comparisons: different split-samples. Normal error distributions with break in variance, with different values of ρ in (2.33) and $\theta = 0.9$ in (2.32)



Note: These figures compare the power envelope the PCC-POS test statistic using different split-samples: 10%, 30%, 50%, 70%. "PE"refers to the power envelope of the PCC-POS test.

to test the null hypothesis (2.21) against the alternative hypothesis (2.22). Formally, this implies finding all the values of $\beta_0 \in \mathbb{R}^k$ such that

$$SN_n(\beta_0 \mid \beta_1) = \sum_{t=2}^{n} \sum_{l=t-1}^{2} \ln c_{\tilde{\sigma}_{lt}t, |\tilde{\sigma}_{1t},\ldots,\tilde{\sigma}_{t-1,t}} + \sum_{t=2}^{n} \ln c_{\tilde{\sigma}_{1t}t} + \sum_{t=1}^{n} s(y_t - \beta_0' x_{t-1}) \tilde{a}_t(\beta_0 \mid \beta_1) < c_1(\beta_0, \beta_1). \tag{2.28}$$

where the critical value is given by the smallest constant $c(\beta_0, \beta_1)$ such that

$$P\left[SN_n(\beta_0 \vert\, \beta_1) > c(\beta_0, \beta_1) \vert \beta = \beta_0\right] \le \alpha$$

Thus, the confidence region of the vector of parameters β can be defined as follows:

$$C_\beta(\alpha) = \left\{\beta_0 : SN_n(\beta_0 \vert\, \beta_1) < c(\beta_0, \beta_1) \vert P[SN_n(\beta_0 \vert\, \beta_1) > c(\beta_0, \beta_1) \vert \beta = \beta_0] \le \alpha\right\}.$$

Given the confidence region $C_\beta(\alpha)$, confidence intervals for the components and sub-vectors of vector β can be derived using the projection techniques. For a review of the projection technique with a numerical illustration, the reader is referred to the first chapter of the thesis [see also Dufour and Taamouti (2010a) and Coudin and Dufour (2009)]. Confidence sets in the form of transformations T of $\beta \in \mathbb{R}^m$, for $m \le (k+1)$, say $T(C_\beta(\alpha))$, can easily be found using these techniques. Since, for any set $C_\beta(\alpha)$

$$\beta \in C_\beta(\alpha) \implies T(\beta) \in T(C_\beta(\alpha)), \tag{2.29}$$

we have

$$P[\beta \in C_\beta(\alpha)] \ge 1 - \alpha \implies P[T(\beta) \in T(C_\beta(\alpha))] \ge 1 - \alpha \tag{2.30}$$

where

$$T(C_\beta(\alpha)) = \{\delta \in \mathbb{R}^m : \exists \beta \in C_\beta(\alpha), T(\beta) = \delta\}.$$

From (2.29) and (2.30) , the set $T(C_\beta(\alpha))$ is a conservative confidence set for $T(\beta)$ with level $1-\alpha$.

If $T(\beta)$ is a scalar, then we have

$$P[\inf\{T(\beta_0),\ \text{ for }\ \beta_0 \in C_\beta(\alpha)\} \leq T(\beta) \leq \sup\{T(\beta_0),\ \text{ for }\ \beta_0 \in C_\beta(\alpha)\}] > 1-\alpha.$$

2.7 Monte Carlo study

In this Section, we assess the performance of the proposed 10% SS-PCC-POS tests (in terms of size control and power) by comparing it with other tests that are intended to be robust against non-standard distributions and heteroskedasticity of unknown form. We consider DGPs under different distributional assumptions and heteroskedasticities. For each DGP, we further consider different correlation coefficients between the residuals of the predictive regressions and the disturbances of the regressors. In the first subsection, the DGPs are formally introduced and in the following subsection, the performance of the proposed 10% SS-PCC-POS tests are compared with that of the other tests considered in our study.

2.7.1 Simulation setup

We assess the performance of the proposed 10% SS-PCC-POS tests in terms of size and power, by considering various DGPs with different symmetric and asymmetric distributions and forms of heteroskedasticity. The DGPs under consideration are supposed to mimic different scenarios that are often encountered in practical settings, the motivation for which have extensively been discussed in the first chapter. The performance of the 10% SS-PCC-POS tests is compared to that of a few other tests, by considering the following linear predictive regression model

$$y_t = \beta x_{t-1} + \varepsilon_t \tag{2.31}$$

where β is an unknown parameter and

$$x_t = \theta x_{t-1} + u_t \tag{2.32}$$

where $\theta = 0.9$ and

$$u_t = \rho\varepsilon_t + w_t\sqrt{1-\rho^2} \tag{2.33}$$

for $\rho = 0, 0.1, 0.5, 0.9$ and ε_t and w_t are assumed to be independent. The initial value of x is given by: $x_0 = \frac{w_0}{\sqrt{1-\theta^2}}$ and w_t are generated from $N(0,1)$. The residuals ε_t are i.n.i.d and are categorized by two groups in our simulation study. In the first group, we consider DGPs where the residuals ε_t possess symmetric and asymmetric distributions:

1. normal distribution: $\varepsilon_t \sim N(0,1)$;

2. Cauchy distribution: $\varepsilon_t \sim Cauchy$;

3. Student's t distribution with two degrees of freedom: $\varepsilon_t \sim t(2)$;

4. Mixture of Cauchy and normal distributions: $\varepsilon_t \sim \mid \varepsilon_t^C \mid -(1-s_t) \mid \varepsilon_t^N \mid$, where ε_t^C follows Cauchy distribution, ε_t^N follows $N(0,1)$ distribution, and

$$P(s_t = 1) = P(s_t = 0) = \frac{1}{2}$$

The second group of DGPs represents different forms of heteroskedasticity:

5. break in variance:

$$\varepsilon_t \sim \begin{cases} N(0,1) & \text{for } t \neq 25 \\ \sqrt{1000}N(0,1) & \text{for } t = 25 \end{cases};$$

6. GARCH(1, 1) plus jump variance:

$$\sigma_\varepsilon^2(t) = 0.00037 + 0.0888\varepsilon_{t-1}^2 + 0.9024\sigma_\varepsilon^2(t-1)\,,$$

$$\varepsilon_t \sim \begin{cases} N(0,\sigma_\varepsilon^2(t)) & \text{for } t \neq 25 \\ 50N(0,\sigma_\varepsilon^2(t)) & \text{for } t = 25 \end{cases};$$

We consider the problem of testing the null hypothesis $H_0 : \beta = 0$. Our Monte Carlo simulations compare the size and power of the 10%-PCC-POS test to those of T-test, T-test based on White

(1980) variance-correction (WT-test hereafter), and the sign-based test proposed by Campbell and Dufour (1995). Due to computational constraints, we perform only $M_1 = 1,000$ simulations to evaluate the probability distribution of the 10% SS-PCC-POS test statistic and $M_2 = 1,000$ iterations for approximating the power functions of the proposed PCC-POS test and other tests. In all simulations, we have considered a sample size of $n = 50$. As the sign-based statistic of Campbell and Dufour (1995) has a discrete distribution, it is not possible to obtain test with a precise size of 5%; therefore, the size of the test is 5.95% for $n = 50$.

2.7.2 Simulation results

The results of the Monte Carlo study corresponding to DGPs described in Section 2.7.1 are presented in figures 2.6-2.11. These figures compare the performance of the 10% SS-PCC-POS test in terms of size and power, to those of the T-test, T-test based on White's (1980) variance-correction, as well as the sign-based procedure proposed by Dufour and Taamouti (2010a). The results are descirbed in detail below.

First, figure 2.6 considers the case where the residuals ε_t are normally distributed. At first glance, we note that all tests control size. Evidently, our test is outperformed by the T-test, as well as the T-test based on White's (1980) variance-correction. The former is expected, since for normally distributed residuals, the T-test is the most powerful test. However, the 10% SS-PCC-POS test outperforms the sign-based procedure porposed by Campbell and Dufour (1995) [CD (1995) hereafter]. Furthermore, changing the correlation coefficient ρ does not seem to lead to visually significant differences in the performance of the tests.

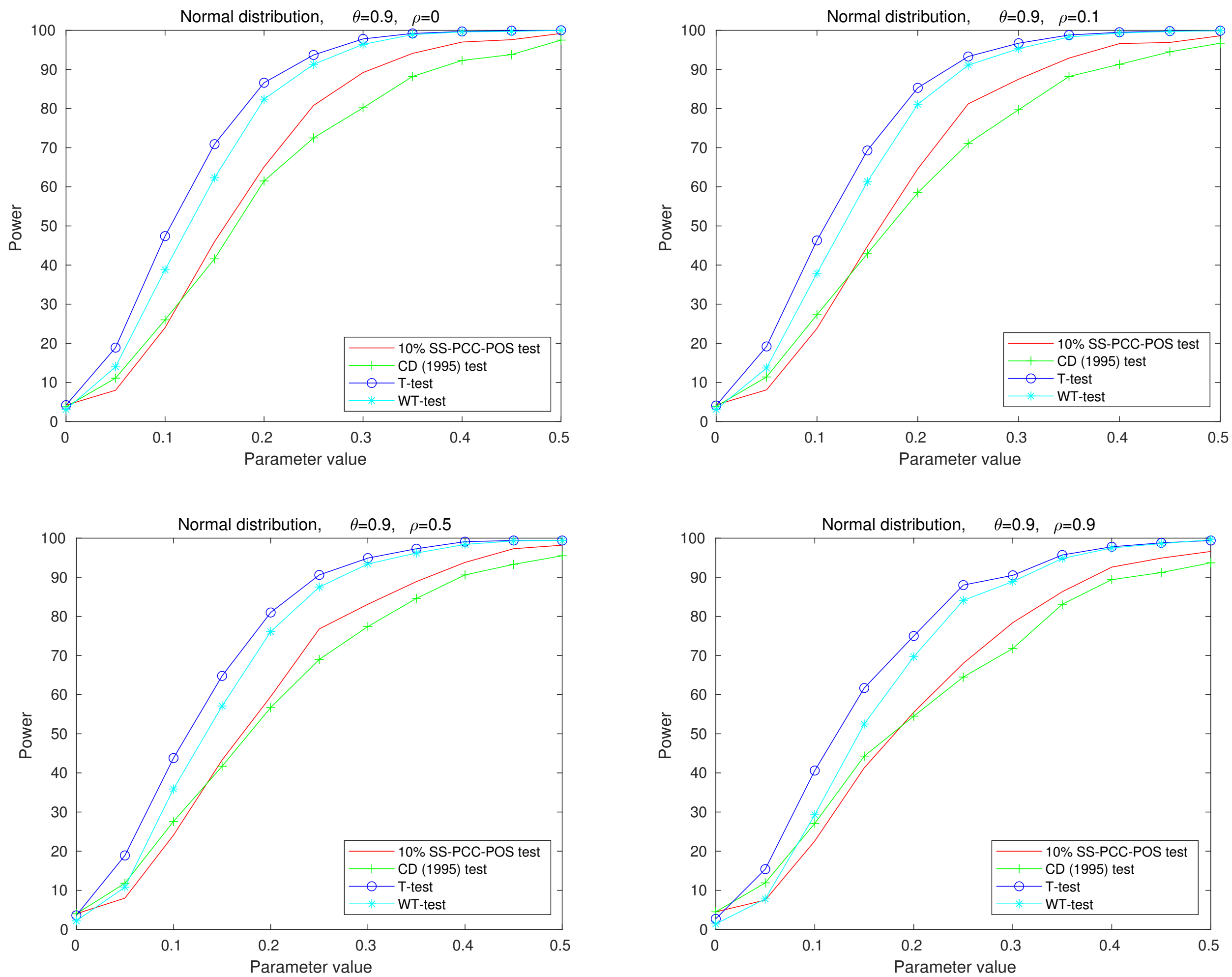
Second, figure 2.7 presents the results of the performance of the aforementioned tests, when the residuals ε_t follows Cauchy distribution. It is evident that the 10% SS-PCC-POS test outperforms all other tests. Moreover, the T-test and WT-test do not possess much power for low correlation coefficient (0 and 0.1) values of ρ. However, as the correlation between u_t and w_t increases, the gap between the power functions narrows significantly.

Third, in figures 2.8 and 2.9, we have considered the cases where the residuals in turn follow $t(2)$ and mixture distributions. In the former case of $t(2)$ distributed residuals, the 10% SS-PCC-POS test outperforms the rest; however, for almost all correlation coefficients ρ, the gap between the power functions is rather small, albeit it is narrowest for $\rho = 0.9$. In the case of residuals with mixture distribution, our 10% SS-PCC-POS test is still the most powerful test. On other hand, it is evident that the T-test and WT-test do not possess much power for small values (0 and 0.1) of correlation coefficient ρ. However, the power functions increase and converge to those of the other tests, as the correlation increases.

An interesting observation is the stark contrast between the power of the 10% SS-PCC-POS test and the T-test, when the errors follow the Cauchy, $t(2)$ and normal distributions respectively. As it has been discussed in the first chapter, the Cauchy and $t(2)$ distributions possess heavy tails, in the presence of which the standard error of the regression coefficients is inflated, which in turn leads to low power when the mean is used as a measure of central tendency. For instance, the Cauchy distribution has the heaviest tails among the considered DGPs, as a result of which the T-test and WT-test have very low power. By noting that a Student's t distribution with ν degrees of freedom converges to the Cauchy distribution for $\nu = 1$ and to the normal distribution as $\nu \to \infty$, it would be interesting to see at which degree of freedom the 10% SS-PCC-POS test is outperformed by the T-test and WT-tests. Figures 2.12-2.15 suggest that for different values of ρ in (2.33) the T-test and WT test outperform the 10% SS-PCC-POS test for $\nu = 4$. Interestingly, figure 2.16 shows that the tails of the $t(2)$ distribution are substantially heavier than that of the $t(4)$, which may explain the transition.

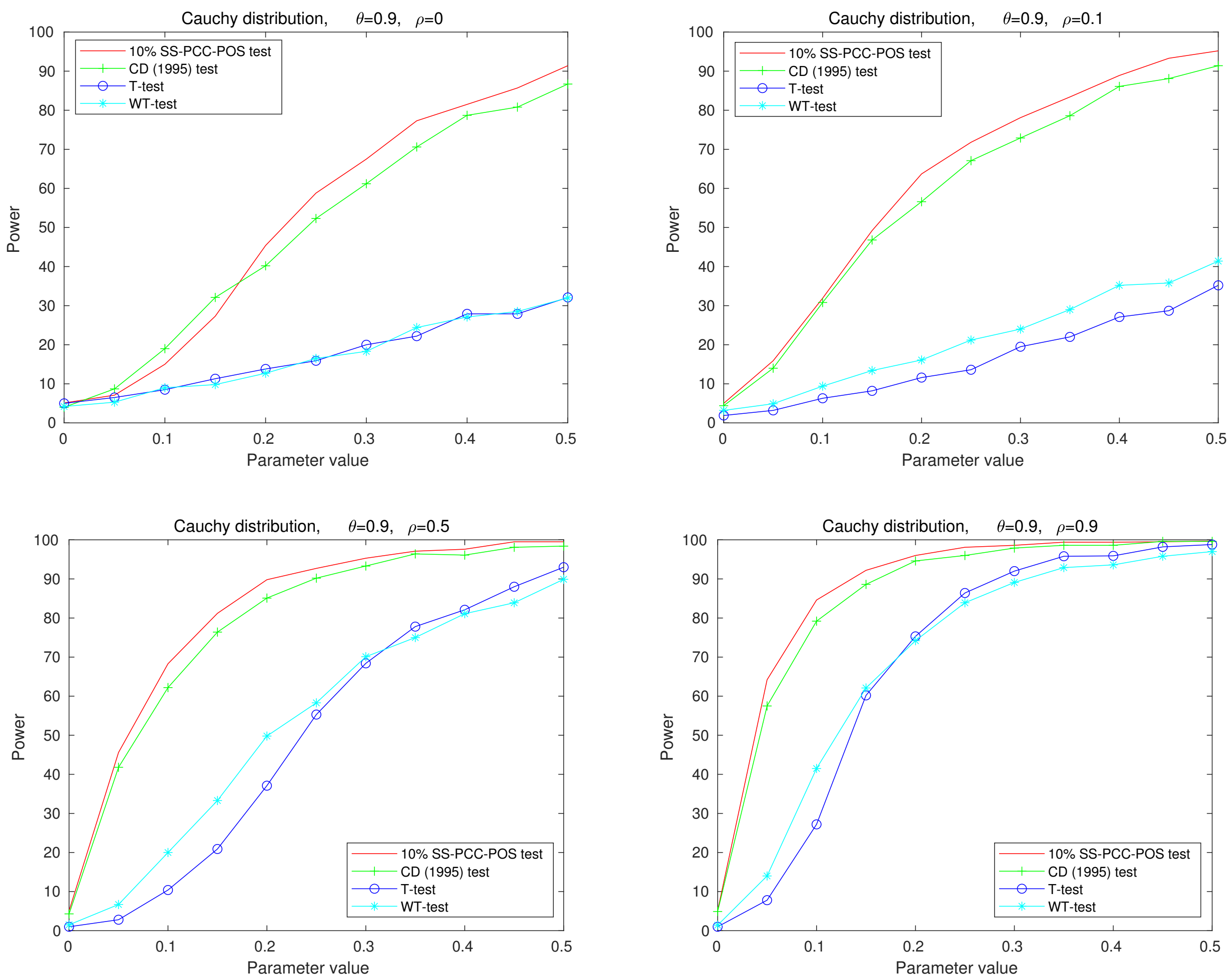
Finally, in figures 2.8 and 2.9, the residuals are normally distributed with different forms of heteroskedasticity. In the first case [see figure 2.8], there is a break in variance, in the presence of which, our test outperforms the other tests. Furthermore, the T-test and WT-test do not possess any power for low correlation (0 and 0.1) values of ρ. However, with greater values of the correlation coefficient the power curves of all test appear to converge. In the other case [see figure 2.9], the variance follows a GARCH(1,1) model with a jump in variance. In this case, our test is only outperformed by the CD (1995) test, which has the greatest power function. Nevertheless, the 10% SS-PCC-POS test still outperforms the T-test and WT-test.

Figure 2.6: Power comparisons: different tests. Normal error distributions with different values of ρ in (2.33) and $\theta = 0.9$ in (2.32)



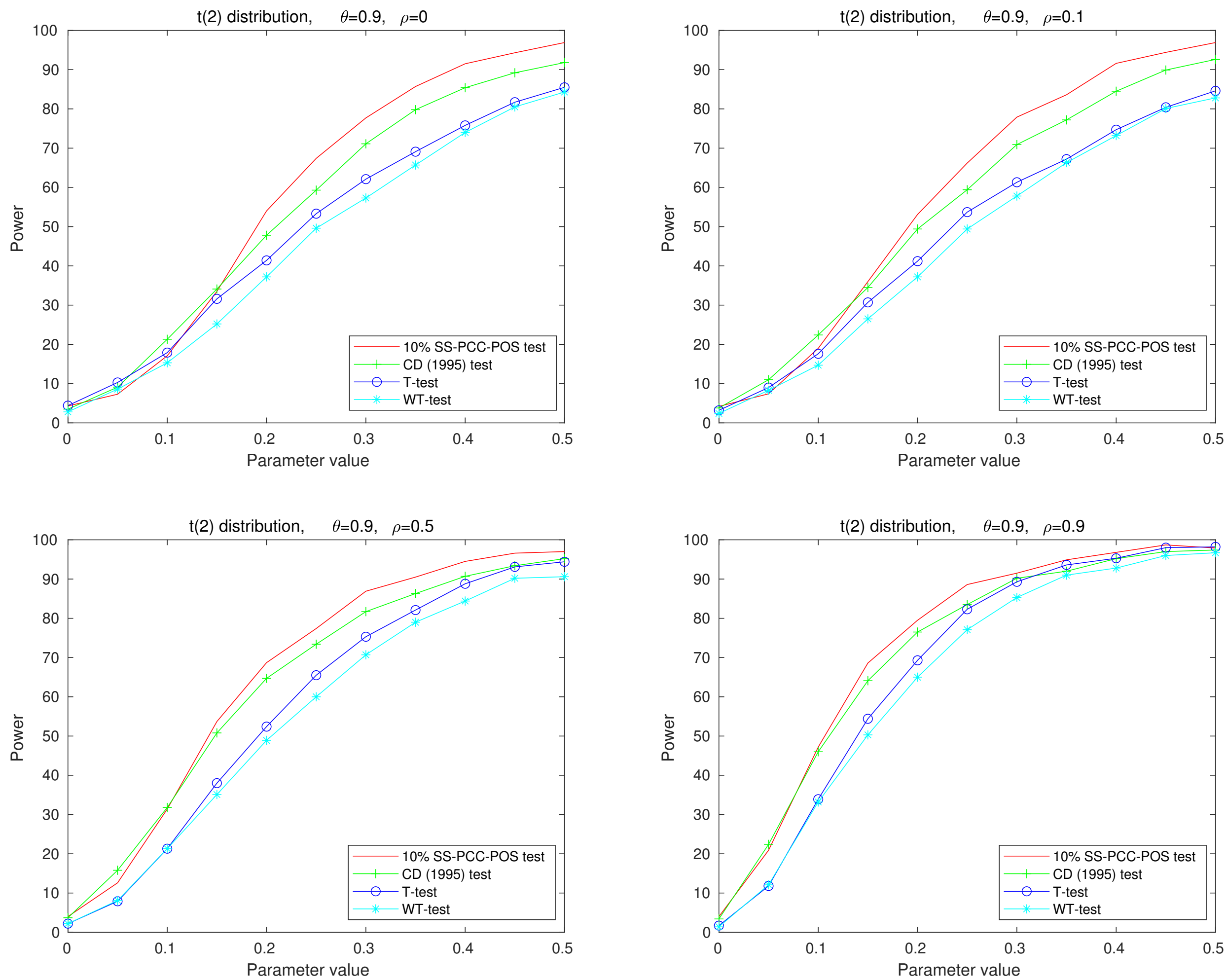
Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

Figure 2.7: Power comparisons: different tests. Cauchy error distributions with different values of ρ in (2.33) and $\theta = 0.9$ in (2.32)



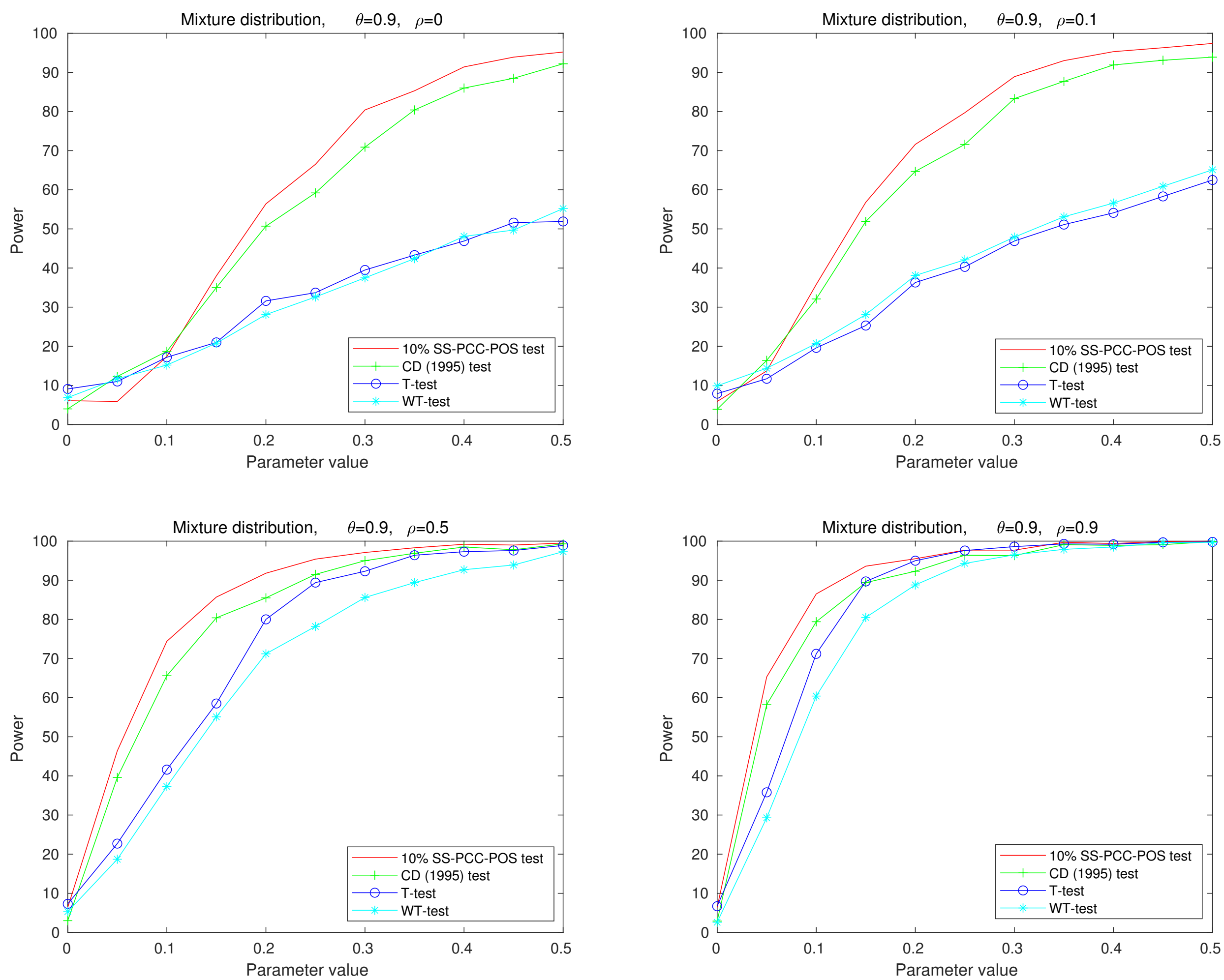
Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

Figure 2.8: Power comparisons: different tests. Student's t error distributions with 2 degrees of freedom [i.e $t(2)$], with different values of ρ in (2.33) and $\theta = 0.9$ in (2.32)



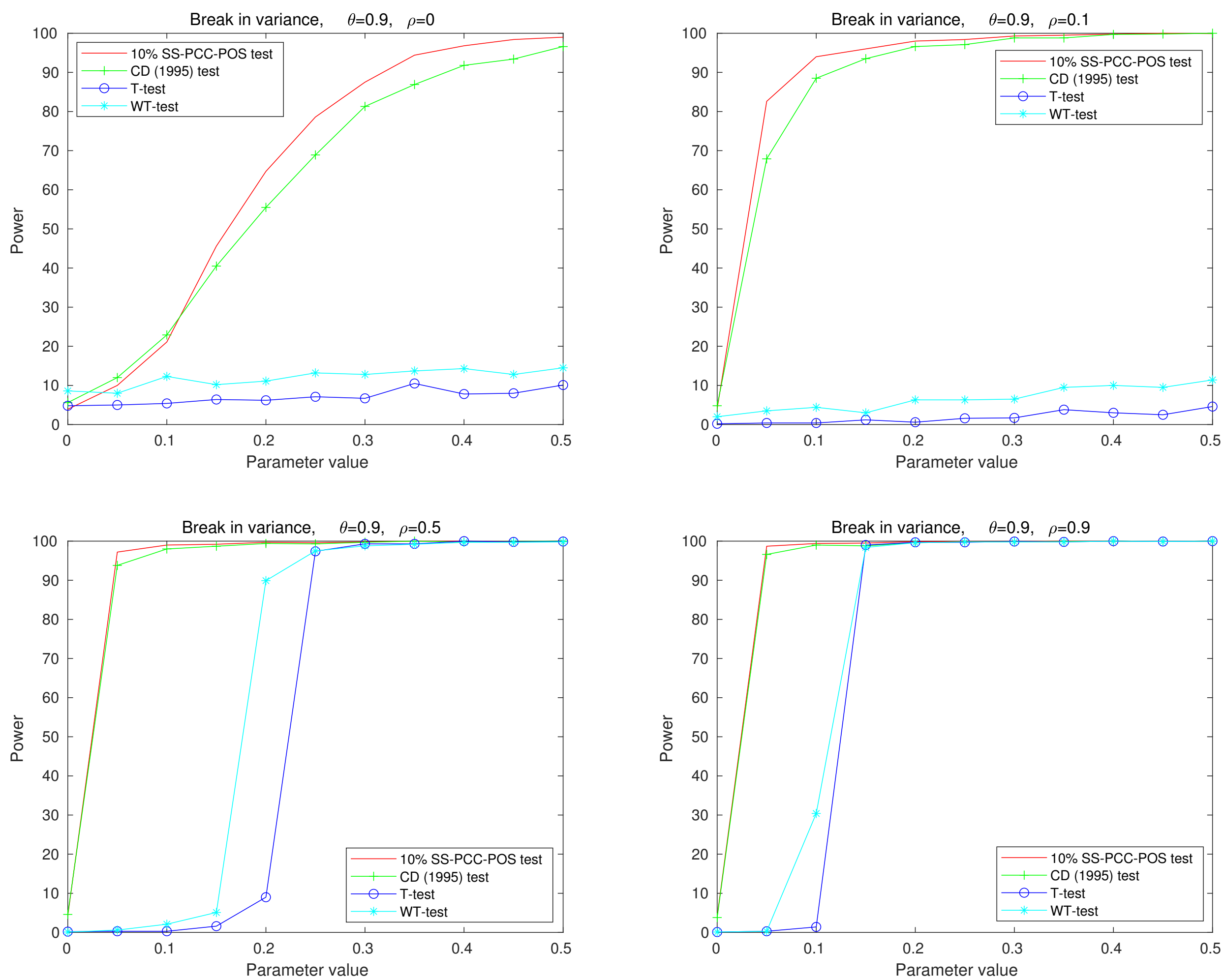
Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

Figure 2.9: Power comparisons: different tests. Mixture error distributions with different values of ρ in (2.33) and $\theta = 0.9$ in (2.32)



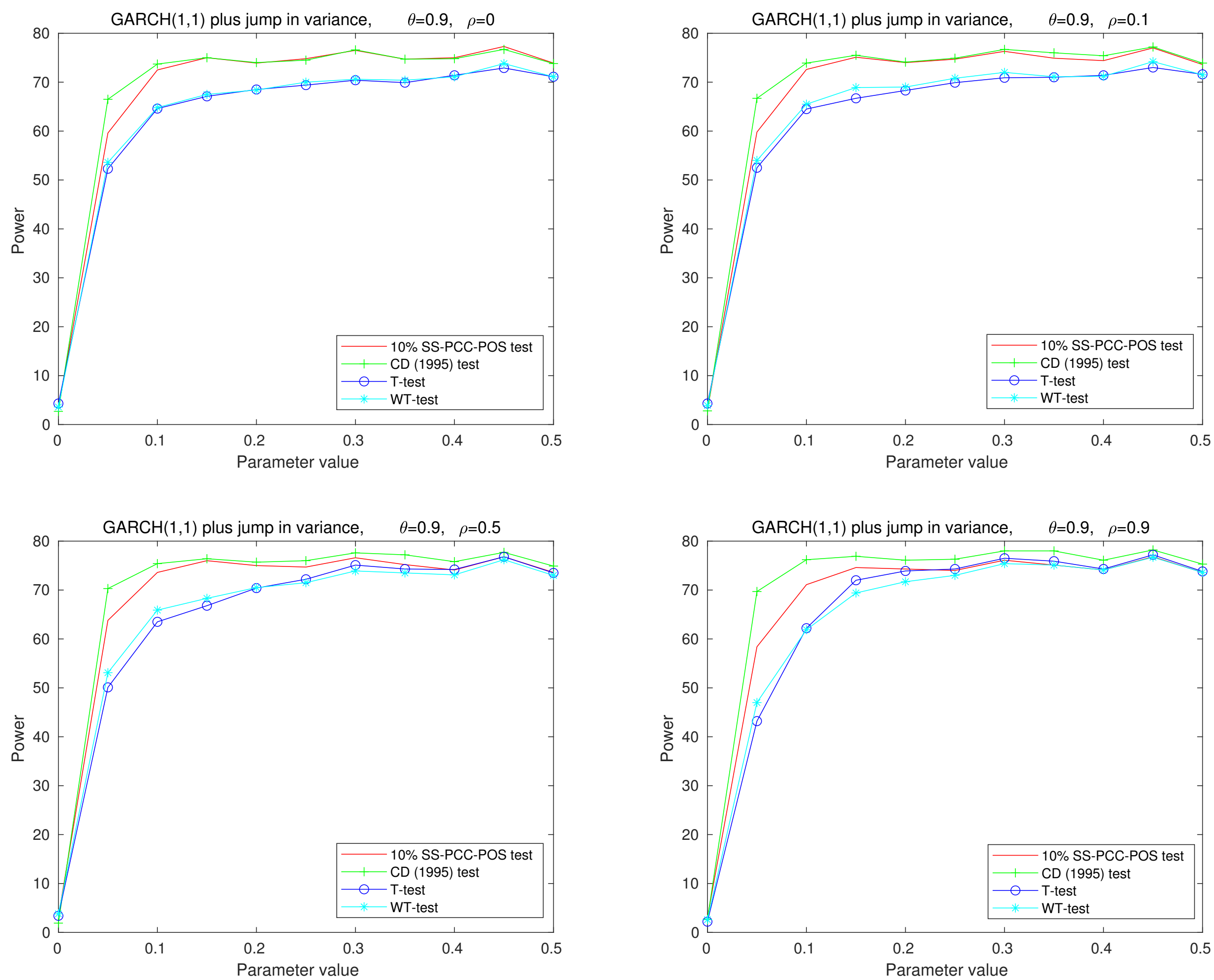
Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

Figure 2.10: Power comparisons: different tests. Normal error distributions with break in variance, with different values of ρ in (2.33) and $\theta = 0.9$ in (2.32)



Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

Figure 2.11: Power comparisons: different tests. Normal error distributions GARCH(1,1) plus jump invariance, with different values of ρ in (2.33) and $\theta = 0.9$ in (2.32)



Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the T-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the T-test based on White's (1980) variance correction [WT-test].

2.8 Conclusion

In this chapter, we extended the exact point-optimal sign-based procedures proposed in the first chapter by relaxing the Markovian assumption. We had earlier provided a caveat that to obtain feasible test statistics, an assumption must be imposed on the dependence structure of process of signs; specifically, we assumed that the signs follow a finite order Markov process. In this chapter, we showed that by implementing the procedures for pair copula constructions of discrete data, we can derive exact and distribution-free sign-based statistics for dependent data in the context of linear and non-linear regressions, without any potentially restrictive assumptions. The proposed tests are valid, distribution-free and robust against heteroskedasticity of unknown form. Furthermore, they may be inverted to produce a confidence region for the vector (sub-vector) of parameters of the regression model.

We further suggested a sequential estimation strategy for the vine PCC and discussed the choice of the copula family. As the proposed sign-statistics depend on the alternative hypothesis, another problem consists of finding an alternative that controls size and maximizes the power. In line with Dufour and Taamouti (2010a), we find that when 10% of sample is used to estimate the alternative and the rest to compute the test-statistic, our procedures have the optimal power and are closest to the power envelope.

Finally, we presented a Monte Carlo study to assess the performance of the proposed tests in terms of size control and power, by comparing it to some other tests that are supposed to be robust against heteroskedasticity. We considered DGPs similar to those in the first chapter and we showed that the 10% split-sample point-optimal sign-test based on pair copula constructions is more superior to the T-test, Campbell and Dufour (1995) sign-based test, and the T-test based on White (1980) variance correction in most cases.

2.9 Appendix

Derivation of the Neyman-Pearson type test-statistic for testing the orthogonality hypothesis for $n \leq 3$. The likelihood function of sample in terms of signs $s(y_1), ..., s(y_n)$ conditional on X is

$$L\left(U(n), \beta\right) = P\left[s(y_1) = s_1, ..., s(y_n) = s_n \mid X\right] = \prod_{t=1}^{n} P\left[s(y_t) = s_t \mid \underline{S}_{t-1} = \underline{s}_{t-1}, X\right],$$

for

$$\underline{S}_0 = \{\emptyset\}, \quad \underline{S}_{t-1} = \{s(y_1), ..., s(y_{t-1})\}, \text{ for } t \geq 2,$$

and

$$P\left[s(y_1) = s_1 \mid \underline{S}_0 = \underline{s}_0, X\right] = P\left[s(y_1) = s_1 \mid X\right],$$

where each s_i, for $1 \leq t \leq n$, takes two possible values 0 and 1. Given model (2.1) and assumption (2.2), under the null hypothesis the signs $s(\varepsilon_t)$, for $1 \leq t \leq n$, are i.i.d conditional on X according to $Bi(1, 0.5)$. Then, the signs $s(y_t)$, for $1 \leq t \leq n$, will also be i.i.d conditional on X with

$$P\left[s(y_t) = 1 \mid X\right] = P\left[s(y_t) = 0 \mid X\right] = \frac{1}{2}, \text{ for } t = 1, ..., n.$$

Consequently, under H_0

$$L_0\left(U(n), 0\right) = \prod_{t=1}^{n} P\left[s(y_t) = s_t \mid X\right] = \left(\frac{1}{2}\right)^n$$

and under H_1 we have

$$L_1\left(U(n), \beta_1\right) = \prod_{t=1}^{n} P\left[s(y_t) = s_t \mid \underline{S}_{t-1} = \underline{s}_{t-1}, X\right]$$

where now, for $t = 1, ..., n$,

$$y_t = \beta_1' x_{t-1} + \varepsilon_t$$

The log-likelihood ratio is given by

$$\ln\left\{\frac{L_1\left(U(n),\beta_1\right)}{L_0\left(U(n),0\right)}\right\} = \sum_{t=1}^{n} \ln\left\{P\left[s(y_t)=s_t \mid \underline{\mathrm{S}}_{t-1}=\underline{\mathrm{s}}_{t-1}, X\right]\right\} - \mathrm{n}\ln\left\{\frac{1}{2}\right\}.$$

According to Neyman-Pearson lemma [see e.g. Lehmann (1959), page 65], the best test to test H_0 against H_1, based on $s(y_1),\ldots,s(y_n)$, rejects H_0 when

$$SL_n(\beta_1) = \ln\left\{\frac{L_1\left(U(n),\beta_1\right)}{L_0\left(U(n),0\right)}\right\} \geq c$$

or when

$$\sum_{t=1}^{n} \ln\left\{P\left[s(y_t)=s_t \mid \underline{\mathrm{S}}_{t-1}=\underline{\mathrm{s}}_{t-1}, X\right]\right\} \geq c_1 \equiv c + n\ln\left(\frac{1}{2}\right),$$

The critical value, say c_1 is given by the smallest constant c_1 such that

$$P\left(\sum_{t=1}^{n} \ln\left\{P\left[s(y_t)=s_t \mid \underline{\mathrm{S}}_{t-1}=\underline{\mathrm{s}}_{t-1}, X\right]\right\} > c_1 \mid H_0\right) \leq \alpha.$$

Let $X=[x_0,\ldots,x_{n-1}]$ be a $n\times(k+1)$ matrix of fixed or stochastic explanatory variables, from (2.10), we get

$$\begin{aligned}
\ln\left\{P[s(y_1)=s_1 \mid \underline{\mathrm{S}}_0=\underline{\mathrm{s}}_0, X]\right\} &= \ln\left\{P[s(y_1)=s_1 \mid X]\right\} \\
&= s(y_1)\ln\left\{\frac{P[y_1\geq 0 \mid X]}{P[y_1<0 \mid X]}\right\} + \ln P[y_1<0 \mid X]
\end{aligned}$$

and for $t=2,\ldots,n$, with $n\leq 3$ we have

$$\begin{aligned}
\sum_{t=2}^{n} \ln P\left[s(y_t)=s_t \mid \underline{\mathrm{S}}_{t-1}=\underline{\mathrm{s}}_{t-1}, X\right] &= \sum_{t=2}^{n} \ln\left(\frac{P[s(y_t)=s_t, s(y_{t-1})=s_{t-1} \mid \underline{\mathrm{S}}_{t-2}=\underline{\mathrm{s}}_{t-2}, X]}{P[s(y_{t-1})=s_{t-1} \mid \underline{\mathrm{S}}_{t-2}=\underline{\mathrm{s}}_{t-2}, X]}\right) \\
&= \sum_{t=2}^{n} \ln\Bigg(\sum_{k_t=0,1}\sum_{k_{t-1}=0,1}(-1)^{k_t+k_{t-1}} \\
&\quad \times\left\{P[s(y_t)\leq s_t-k_t, s(y_{t-1})\leq s_{t-1}-k_{t-1} \mid \underline{\mathrm{S}}_{t-2}=\underline{\mathrm{s}}_{t-2}, X]\right\} \\
&\quad /P[s(y_{t-1})=s_{t-1} \mid \underline{\mathrm{S}}_{t-2}=\underline{\mathrm{s}}_{t-2}, X]\Bigg)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{t=2}^{n} \ln \Bigg(\sum_{k_t=0,1} \sum_{k_{t-1}=0,1} (-1)^{k_t+k_{t-1}} \\
&\quad \times \Big\{ C_{s(y_t),s(y_{t-1})|\underset{\sim}{\mathrm{S}}_{t-2}} \big(F_{s(y_t)|\underset{\sim}{\mathrm{S}}_{t-2}}(s_t - k_t \mid \underset{\sim}{\mathrm{s}}_{t-2}, X), \\
&\qquad\qquad F_{s(y_{t-1})|\underset{\sim}{\mathrm{S}}_{t-2}}(s_{t-1} - k_{t-1} \mid \underset{\sim}{\mathrm{s}}_{t-2}, X) \big) \Big\} \\
&\quad / P[s(y_{t-1}) = s_{t-1} \mid \underset{\sim}{\mathrm{S}}_{t-2} = \underset{\sim}{\mathrm{s}}_{t-2}, X] \Bigg) \\
&= \sum_{t=2}^{n} \ln \Bigg\{ \sum_{k_t=0,1} \sum_{k_{t-1}=0,1} (-1)^{k_t+k_{t-1}} \\
&\quad \times \Big\{ C_{s(y_t),s(y_{t-1})|\underset{\sim}{\mathrm{S}}_{t-2}} \big(F_{s(y_t)|\underset{\sim}{\mathrm{S}}_{t-2}}(s_t - k_t \mid \underset{\sim}{\mathrm{s}}_{t-2}, X), \\
&\qquad\qquad F_{s(y_{t-1})|\underset{\sim}{\mathrm{S}}_{t-2}}(s_{t-1} - k_{t-1} \mid \underset{\sim}{\mathrm{s}}_{t-2}, X) \big) \Big\} \Bigg\} \\
&\quad - \sum_{t=2}^{n} \ln \left\{ P[s(y_{t-1}) = s_{t-1} \mid \underset{\sim}{\mathrm{S}}_{t-2} = \underset{\sim}{\mathrm{s}}_{t-2}, X] \right\}
\end{aligned}
$$

Each argument $F_{s(y_t)|\underset{\sim}{\mathrm{S}}_{t-2}}(s_t - k_t \mid \underset{\sim}{\mathrm{s}}_{t-2}, X)$ and $F_{s(y_{t-1})|\underset{\sim}{\mathrm{S}}_{t-2}}(s_{t-1} - k_{t-1} \mid \underset{\sim}{\mathrm{s}}_{t-2}, X)$ in the copula expression above can be evaluated as follows

$$
\begin{aligned}
&F_{s(y_t)|\underset{\sim}{\mathrm{S}}_{t-2}}(s_t - k_t \mid \underset{\sim}{\mathrm{s}}_{t-2}, X) = \\
&\big\{ C_{s(y_t),s(y_{t-2})|\underset{\sim}{\mathrm{S}}_{t-3}}(F(s_t - k_t \mid \underset{\sim}{\mathrm{s}}_{t-3}, X), F(s_{t-2} \mid \underset{\sim}{\mathrm{s}}_{t-3}, X)) \\
&- C_{s(y_t),s(y_{t-2})|\underset{\sim}{\mathrm{S}}_{t-3}}(F(s_t - k_t \mid \underset{\sim}{\mathrm{s}}_{t-3}, X), F(s_{t-2} - 1 \mid \underset{\sim}{\mathrm{s}}_{t-3}, X)) \big\} / P[s(y_{t-2}) = s_{t-2} \mid \underset{\sim}{\mathrm{S}}_{t-3} = \underset{\sim}{\mathrm{s}}_{t-3}, X]
\end{aligned}
$$

and similarly

$$
\begin{aligned}
&F_{s(y_{t-1})|\underset{\sim}{\mathrm{S}}_{t-2}}(s_{t-1} - k_{t-1} \mid \underset{\sim}{\mathrm{s}}_{t-2}, X) = \\
&\big\{ C_{s(y_{t-2}),s(y_{t-1})|\underset{\sim}{\mathrm{S}}_{t-3}}(F(s_{t-2} \mid \underset{\sim}{\mathrm{s}}_{t-3}, X), F(s_{t-1} - k_{t-1} \mid \underset{\sim}{\mathrm{s}}_{t-3}, X)) \\
&- C_{s(y_{t-2}),s(y_{t-1})|\underset{\sim}{\mathrm{S}}_{t-3}}(F(s_{t-2} - 1 \mid \underset{\sim}{\mathrm{s}}_{t-3}, X), F(s_{t-1} - k_{t-1} \mid \underset{\sim}{\mathrm{s}}_{t-3}, X)) \big\} / P[s(y_{t-2}) = s_{t-2} \mid \underset{\sim}{\mathrm{S}}_{t-3} = \underset{\sim}{\mathrm{s}}_{t-3}, X]
\end{aligned}
$$

Thus, for $n \leq 3$ the Neyman-Pearson type test statistic based on $s(y_1), ..., s(y_n)$, can be expressed as

$$\ln\left\{\frac{L_1\left(U(n),\beta_1\right)}{L_0\left(U(n),0\right)}\right\} = \ln P[s(y_1)=s_1 \mid \underset{\sim}{\mathrm{S}}_0 = \underset{\sim}{\mathrm{s}}_0, X] + \sum_{t=2}^{n} \ln\left\{\sum_{k_t=0,1}\sum_{k_{t-1}=0,1}(-1)^{k_t+k_{t-1}}\right.$$
$$\left.\times\left(C_{s(y_t),s(y_{t-1})|\underset{\sim}{\mathrm{S}}_{t-2}}\left(F_{s(y_t)|\underset{\sim}{\mathrm{S}}_{t-2}}(s_t-k_t \mid \underset{\sim}{\mathrm{s}}_{t-2},X), F_{s(y_{t-1})|\underset{\sim}{\mathrm{S}}_{t-2}}(s_{t-1}-k_{t-1} \mid \underset{\sim}{\mathrm{s}}_{t-2},X)\right)\right)\right\}$$
$$-\sum_{t=2}^{n}\ln\left\{P[s(y_{t-1})=s_{t-1} \mid \underset{\sim}{\mathrm{S}}_{t-2}=\underset{\sim}{\mathrm{s}}_{t-2},X]\right\} - n\ln\left\{\frac{1}{2}\right\}$$

■

Vine decomposition in the continuous case. In Section 2.4, it has been shown that the signs $s(y_1), ..., s(y_n)$ may have a continuous extension with a perturbation in $[0,1]$ [see Denuit and Lambert (2005)]. This can be achieved by employing a transformation of the form $s^*(y_t) = s(y_t) + U - 1$ for $t = 1, ..., n$, where a natural choice for U is the uniform distribution. Thus, for $\{s^*(y_t) \in \mathbb{R}, t = 1, ..., n\}$ consider the continuous equivalent of the conditional probability mass function (2.11) - i.e. the conditional density function. Further, by letting $\underset{\sim}{\mathrm{S}}^*_{t-1}$ be the continuous extension of $\underset{\sim}{\mathrm{S}}_{t-1}$, the conditional density function may be expressed as

$$f_{s^*(y_t)|\underset{\sim}{\mathrm{S}}^{*\backslash j}_{t-1}\cup s^*(y_j)} = \frac{f_{s^*(y_t),s^*(y_j)|\underset{\sim}{\mathrm{S}}^{*\backslash j}_{t-1}}}{f_{s^*(y_j)|\underset{\sim}{\mathrm{S}}^{*\backslash j}_{t-1}}}. \tag{2.34}$$

From the Theorem of Sklar (1959), we know that

$$f_{s^*(y_t),s^*(y_j)|\underset{\sim}{\mathrm{S}}^{*\backslash j}_{t-1}}(s^*_t, s^*_j \mid \underset{\sim}{\mathrm{s}}^{*\backslash j}_{t-1}, X) = c_{s^*(y_t),s^*(y_j)|\underset{\sim}{\mathrm{S}}^{*\backslash j}_{t-1}}\left(F_{s^*(y_t)|\underset{\sim}{\mathrm{S}}^{*\backslash j}_{t-1}}(s^*_t \mid \underset{\sim}{\mathrm{s}}^{*\backslash j}_{t-1}, X), F_{s^*(y_j)|\underset{\sim}{\mathrm{S}}^{*\backslash j}_{t-1}}(s^*_j \mid \underset{\sim}{\mathrm{s}}^{*\backslash j}_{t-1}, X)\right)$$
$$\times \quad f_{s^*(y_t)|\underset{\sim}{\mathrm{S}}^{*\backslash j}_{t-1}} f_{s^*(y_j)|\underset{\sim}{\mathrm{S}}^{*\backslash j}_{t-1}}, \tag{2.35}$$

where $c()$ is the copula density function. Thus,

$$f_{s^*(y_t)|\underset{\sim}{\mathrm{S}}^{*\backslash j}_{t-1}\cup s^*(y_j)} = c_{s^*(y_t),s^*(y_j)|\underset{\sim}{\mathrm{S}}^{*\backslash j}_{t-1}}\left(F_{s^*(y_t)|\underset{\sim}{\mathrm{s}}^{*\backslash j}_{t-1}}(s^*_t \mid \underset{\sim}{\mathrm{s}}^{*\backslash j}_{t-1}, X), F_{s^*(y_j)|\underset{\sim}{\mathrm{S}}^{*\backslash j}_{t-1}}(s^*_j \mid \underset{\sim}{\mathrm{s}}^{*\backslash j}_{t-1}, X)\right) f_{s^*(y_t)|\underset{\sim}{\mathrm{S}}^{*\backslash j}_{t-1}}, \tag{2.36}$$

with

$$c_{s^*(y_t),s^*(y_j)|\underline{\mathbb{S}}^{*\setminus j}_{t-1}}\left(F_{s^*(y_t)|\underline{\mathbb{S}}^{*\setminus j}_{t-1}}(s^*_t \mid \underline{\mathrm{s}}^{*\setminus j}_{t-1}, X), F_{s^*(y_j)|\underline{\mathbb{S}}^{*\setminus j}_{t-1}}(s^*_j \mid \underline{\mathrm{s}}^{*\setminus j}_{t-1}, X)\right) = \frac{\partial^2 C_{s^*(y_t),s^*(y_j)|\underline{\mathbb{S}}^{*\setminus j}_{t-1}}\left(F_{s^*(y_t)|\underline{\mathbb{S}}^{*\setminus j}_{t-1}}(s^*_t|\underline{\mathrm{s}}^{*\setminus j}_{t-1},X),F_{s^*(y_j)|\underline{\mathbb{S}}^{*\setminus j}_{t-1}}(s^*_j|\underline{\mathrm{s}}^{*\setminus j}_{t-1},X)\right)}{\partial F_{s^*(y_t)|\underline{\mathbb{S}}^{*\setminus j}_{t-1}}\left(s^*_t|\underline{\mathrm{s}}^{*\setminus j}_{t-1},X\right)\partial F_{s^*(y_j)|\underline{\mathbb{S}}^{*\setminus j}_{t-1}}\left(s^*_j|\underline{\mathrm{s}}^{*\setminus j}_{t-1},X\right)} \tag{2.37}$$

can express (2.34), and the arguments of the copulae, say, $F_{s^*(y_t)|\underline{\mathbb{S}}^{*\setminus j}_{t-1}}(s^*_t \mid \underline{\mathrm{s}}^{*\setminus j}_{t-1}, X)$ are obtained using the expression by Joe (1996), such that

$$F_{s^*(y_t)|\underline{\mathbb{S}}^{*\setminus j}_{t-1}}(s^*_t \mid \underline{\mathrm{s}}^{*\setminus j}_{t-1}, X) = \frac{\partial C_{s^*(y_t),s^*(y_i)|\underline{\mathbb{S}}^{*\setminus j,i}_{t-1}}\left(F_{s^*(y_t)|\underline{\mathbb{S}}^{*\setminus j,i}_{t-1}}(s^*_t \mid \underline{\mathrm{s}}^{*\setminus j,i}_{t-1}, X), F_{s^*(y_i)|\underline{\mathbb{S}}^{*\setminus j,i}_{t-1}}(s^*_i \mid \underline{\mathrm{s}}^{*\setminus j,i}_{t-1}, X)\right)}{\partial F_{s^*(y_i)|\underline{\mathbb{S}}^{*\setminus j,i}_{t-1}}(s^*_i \mid \underline{\mathrm{s}}^{*\setminus j,i}_{t-1}, X)}. \tag{2.38}$$

Therefore, when the data is continuous, the marginals in the copula expressions of, say, the third tree, $F_{t|t+1,t+2}$ for $t = 1, ..., n-2$ and $F_{t+3|t+1,t+2}$ for $t = 1, ..., n-3$ are obtained by

$$F_{t|t+1,t+2} = \frac{\partial C_{t,t+1|t+2}(F_{t|t+2}(s^*_t \mid s^*_{t+2}, X), F_{t+1|t+2}(s^*_{t+1} \mid s^*_{t+2}, X))}{\partial F_{t+1|t+2}(s^*_{t+1} \mid s^*_{t+2}, X)}, \tag{2.39}$$

where $F_{t+3|t+1,t+2}$ is obtained in a similar way. ■

Proof of Proposition 3. The likelihood function of sample in terms of signs $s(y_1), ..., s(y_n)$

$$L\left(U(n), \beta\right) = P\left[s(y_1) = s_1, ..., s(y_n) = s_n \mid X\right]$$

where each s_i, for $1 \leq t \leq n$, takes two possible values 0 and 1. Given model (2.1) and assumption (2.2), under the null hypothesis the signs $s(\varepsilon_t)$, for $1 \leq t \leq n$, are i.i.d conditional on X according to $Bi(1, 0.5)$. Then, the signs $s(y_t)$, for $1 \leq t \leq n$, will also be i.i.d conditional on X

$$P\left[s(y_t) = 1 \mid X\right] = P\left[s(y_t) = 0 \mid X\right] = \frac{1}{2}, \text{ for } t = 1, ..., n.$$

Consequently, under H_0 we have

$$L_0\left(U(n),0\right)=\prod_{t=1}^{n}P\left(s(y_t)=s_t\mid X\right)=\left(\frac{1}{2}\right)^{n}$$

and under H_1 the likelihood function can be expressed as

$$L_1\left(U(n),\beta_1\right)=P_1[s(y_1)=s_1\mid X]\times\prod_{t=2}^{n}P_{t|1:t-1}[s(y_t)=s_t\mid s(y_1)=s_1:s(y_{t-1})=s_{t-1},X].$$

which can further be decomposed using the D-vine array $A=(\sigma_{lt})_{1\leq l\leq t\leq n}$ to obtain

$$L_1\left(U(n),\beta_1\right)=P_1[s(y_1)=s_1\mid X]\times\prod_{t=2}^{n}\prod_{l=t-1}^{2}c_{\sigma_{lt}t,|\sigma_{1t},\ldots,\sigma_{t-1,t}}\times c_{\sigma_{1t}t}\times P_t[s(y_t)=s_t\mid X]$$

where now for $t=1,...,n$,

$$y_t=\beta_1'x_{t-1}+\varepsilon_t$$

Under assumption (2.1) and (2.2), the likelihood function under the alternative can be expressed as

$$\begin{aligned}L_1\left(U(n),\beta_1\right)\;=\;&(1-P_1[\varepsilon_1<-\beta_1x_0\mid X])^{s(y_1)}\times P_1[\varepsilon_1<-\beta_1x_0\mid X]^{1-s(y_1)}\\&\times\prod_{t=2}^{n}\prod_{l=t-1}^{2}c_{\sigma_{lt}j,|\sigma_{1t},\ldots,\sigma_{t-1,t}}\times c_{\sigma_{1t}t}\times(1-P_t[\varepsilon_t<-\beta_1x_{t-1}\mid X])^{s(y_t)}\\&\times P_t[\varepsilon_t<-\beta_1x_{t-1}\mid X]^{1-s(y_t)}\end{aligned}$$

The log-likelihood ratio is given by

$$\begin{aligned}\ln\left\{\frac{L_1\left(U(n),\beta_1\right)}{L_0\left(U(n),0\right)}\right\}\;=\;&s(y_1)\ln\left\{\frac{1-P_1[\varepsilon_1<-\beta_1x_0\mid X]}{P_1[\varepsilon_1<-\beta_1x_0\mid X]}\right\}+\ln\left(1-P_1[\varepsilon_1<-\beta_1x_0\mid X]\right)\\&+\sum_{t=2}^{n}\sum_{l=t-1}^{2}\ln c_{\sigma_{lt}t,|\sigma_{1t},\ldots,\sigma_{t-1,t}}+\sum_{t=2}^{n}\ln c_{a_{1t}t}+\sum_{t=2}^{n}s(y_t)\ln\left\{\frac{1-P_t[\varepsilon_t<-\beta_1x_{t-1}\mid X]}{P_t[\varepsilon_t<-\beta_1x_{t-1}\mid X]}\right\}\\&+\sum_{t=2}^{n}\ln\left(1-P_t[\varepsilon_t<-\beta_1x_{t-1}\mid X]\right)-n\ln\left(\frac{1}{2}\right)\end{aligned}$$

According to Neyman-Pearson Lemma [see e.g. Lehmann (1959), page 65], the best test to test

H_0 against H_1, based on $s(y_1), ..., s(y_n)$, rejects H_0 when

$$\ln\left\{\frac{L_1\left(U(n),\beta_1\right)}{L_0\left(U(n),0\right)}\right\} \geq c$$

or when

$$\begin{aligned}\ln\left\{\frac{L_1\left(U(n),\beta_1\right)}{L_0\left(U(n),0\right)}\right\} &= \sum_{t=2}^{n}\sum_{l=t-1}^{2}\ln c_{\sigma_{lt}t,|\sigma_{1t},\dots,\sigma_{t-1,t}} + \sum_{t=2}^{n}\ln c_{\sigma_{1t}t} \\ &+\sum_{t=1}^{n} s_t \ln\left\{\frac{1-P_t[\varepsilon_t < -\beta_1 x_{t-1} \mid X]}{P_t[\varepsilon_t < -\beta_1 x_{t-1} \mid X]}\right\} > c_1(\beta_1)\end{aligned}$$

The critical value, say $c_1(\beta_1)$ is given by the smallest constant $c_1(\beta_1)$ such that

$$P\left(\ln\left\{\frac{L_1\left(U(n),\beta_1\right)}{L_0\left(U(n),0\right)}\right\} > c_1(\beta_1) \mid H_0\right) \leq \alpha.$$

■ **Algorithm for the likelihood function of the signs under the alternative hypothesis.** In this Section, we adapt the algorithm for the joint pmf for D-vine for discrete variables of Panagiotelis et al. (2012) and Joe (2014) to the context of our study. Let $U(n) = (s(y_1), s(y_2), ..., s(y_n))'$ be a binary valued n-vector. Furthermore, for a vector of integers $\mathbf{i}$, let $\mathbf{S_i} = \{s(y_i), i \in \mathbf{i}\}$, where $\mathbf{s_i}$ is a mass point of $\mathbf{S_i}$ and s_g is a mass point of $s(y_g)$. Let

$$\begin{aligned}F^{+}_{g|\mathbf{i}} &:= P\left[s(y_g) \leq s_g \mid \mathbf{S_i} = \mathbf{s_i}, X\right], \quad F^{-}_{g|\mathbf{i}} := P\left[s(y_g) < s_g \mid \mathbf{S_i} = \mathbf{s_i}, X\right], \\ f_{g|\mathbf{i}} &:= P[s(y_g) = s_g \mid \mathbf{S_i} = \mathbf{s_i}, X].\end{aligned}$$

noting that when $\mathbf{i} = \{\emptyset\}$, these conditional probabilities, correspond to marginal probabilities. Furthermore, let $C_{gh|\mathbf{i}}$ be a bivariate copula for the conditional CDFs $F_{g|\mathbf{i}}$ and $F_{h|\mathbf{i}}$, and denote

$$\begin{aligned}C^{++}_{gh|\mathbf{i}} &:= C_{gh|\mathbf{i}}\left(F^{+}_{g|\mathbf{i}}, F^{+}_{h|\mathbf{i}}\right), \quad C^{+-}_{gh|\mathbf{i}} := C_{gh|\mathbf{i}}\left(F^{+}_{g|\mathbf{i}}, F^{-}_{h|\mathbf{i}}\right), \\ C^{-+}_{gh|\mathbf{i}} &:= C_{gh|\mathbf{i}}\left(F^{-}_{g|\mathbf{i}}, F^{+}_{h|\mathbf{i}}\right), \quad C^{--}_{gh|\mathbf{i}} := C_{gh|\mathbf{i}}\left(F^{-}_{g|\mathbf{i}}, F^{-}_{h|\mathbf{i}}\right).\end{aligned}$$

The main elements of the algorithm is the following recursions:

(I) $F^{+}_{j-t|(j-t+1):(j-1)} = \left[C^{++}_{j-t,j-1|(j-t+1):(j-2)} - C^{+-}_{j-t,j-1|(j-t+1):(j-2)}\right] / f_{j-1|(j-t+1):(j-2)}$;

(II) $F^{-}_{j-t|(j-t+1):(j-1)} = \left[C^{-+}_{j-t,j-1|(j-t+1):(j-2)} - C^{--}_{j-t,j-1|(j-t+1):(j-2)}\right] / f_{j-1|(j-t+1):(j-2)}$;

(III) $f_{j-t|(j-t+1):(j-1)} = F^{+}_{j-t|(j-t+1):(j-1)} - F^{-}_{j-t|(j-t+1):(j-1)}$;

(IV) $F^{+}_{j|(j-t+1):(j-1)} = \left[C^{++}_{j-t+1,j|(j-t+2):(j-1)} - C^{-+}_{j-t+1,j|(j-t+2):(j-1)}\right] / f_{j-t+1|(j-t+2):(j-1)}$;

(V) $F^{-}_{j|(j-t+1):(j-1)} = \left[C^{+-}_{j-t+1,j|(j-t+2):(j-1)} - C^{--}_{j-t+1,j|(j-t+2):(j-1)}\right] / f_{j-t+1|(j-t+2):(j-1)}$;

(VI) $f_{j|(j-t+1):(j-1)} = F^{+}_{j|(j-t+1):(j-1)} - F^{-}_{j|(j-t+1):(j-1)}$;

(VII) The values based on $C_{j-t,j|(j-t+1):(j-1)}$ is computed;

(VIII) t is incremented by 1 and back to (I).

The identity employed in the recursions is

$$P\left[s(y_g) \leq s_g \mid s(y_h) = s_h, \mathbf{S_i} = \mathbf{s_i}, X\right] = \frac{P[s(y_g) \leq s_g, s(y_h) \leq s_h | \mathbf{S_i} = \mathbf{s_i}, X] - P[s(y_g) \leq s_g, s(y_h) < s_h | \mathbf{S_i} = \mathbf{s_i}, X]}{P[s(y_h) = s_h | \mathbf{S_i} = \mathbf{s_i}, X]}.$$

The algorithm is as follows

1. Input $\mathbf{s}_n = (s_1, ..., s_n)$.
2. Allocate an $n \times n$ matrix π, where $\pi_{tj} = f_{(j-t+1):j}$ for $t = 1, ..., n$ and $j = t+1, ..., n$ and the likelihood function $P[s(y_1) = s_1, ..., s(y_n) = s_n]$ under the alternative will appear as π_{nn}.
3. Allocate C^{++}, C^{+-}, C^{-+}, C^{--}, U'^{+}, U'^{-}, U^{+}, U^{-}, u', u, w', w, as vectors of length n.
4. Evaluate F^{+}_{j}, F^{-}_{j}, and $f_j = F^{+}_{j} - F^{-}_{j}$ using (2.14) and let $\pi_{1j} \leftarrow f_j$ for $j = 1, ..., n$;
5. Let $C^{++}_{j} \leftarrow C_{j-1,j}\left(F^{+}_{j-1}, F^{+}_{j}\right)$, $C^{+-}_{j} \leftarrow C_{j-1,j}\left(F^{+}_{j-1}, F^{-}_{j}\right)$, $C^{-+}_{j} \leftarrow C_{j-1,j}\left(F^{-}_{j-1}, F^{+}_{j}\right)$, and $C^{--}_{j} \leftarrow C_{j-1,j}\left(F^{-}_{j-1}, F^{-}_{j}\right)$ for $j = 2, ..., n$;
6. Set $P_{2j} \leftarrow C^{++}_{j} - C^{+-}_{j} - C^{-+}_{j} + C^{--}_{j}$ for $j = 2, ..., n$;
7. **for** $j = 2, ..., n : (T_1)$ **do**

8. $U_j^{\prime +} \leftarrow F_{j-1|j}^{+} = \left(C_j^{++} - C_j^{+-}\right)/f_j$, $U_j^{\prime -} \leftarrow F_{j-1|j}^{-} = \left(C_j^{-+} - C_j^{--}\right)/f_j$, and $u_j^{\prime} \leftarrow f_{j-1|j} = F_{j-1|j}^{+} - F_{j-1|j}^{-}$;
9. $U_j^{+} \leftarrow F_{j|j-1}^{+} = \left(C_j^{++} - C_j^{-+}\right)/f_{j-1}$, $U_j^{-} \leftarrow F_{j|j-1}^{-} = \left(C_j^{-+} - C_j^{--}\right)/f_{j-1}$, and $u_j \leftarrow f_{j|j-1} = F_{j|j-1}^{+} - F_{j|j-1}^{-}$;
10. **end for**
11. **for** $t = 2, ..., n-1 : (T_2, ..., T_{n-1})$ **do**
12. let $C_j^{\alpha\beta} \leftarrow C_{j-t,j|(j-t+1):(j-1)}\left(U_{j-1}^{\prime\alpha}, U_j^{\beta}\right)$, for $j = t+1, ..., n$ and $\alpha, \beta \in \{+, -\}$;
13. let $w_j^{\prime} \leftarrow u_j^{\prime}$, $w_j \leftarrow u_j$ for $j = t, ..., n$;
14. **for** $j = t+1, ..., n$: **do**
15. $U_j^{\prime +} \leftarrow \left(C_j^{++} - C_j^{+-}\right)/w_j$, $U_j^{\prime -} \leftarrow \left(C_j^{-+} - C_j^{--}\right)/w_j$ and $u_j^{\prime} \leftarrow U_j^{\prime +} - U_j^{\prime -}$;
16. $U_j^{+} \leftarrow \left(C_j^{++} - C_j^{-+}\right)/w_{j-1}^{\prime}$, $U_j^{-} \leftarrow \left(C_j^{+-} - C_j^{--}\right)/w_{j-1}^{\prime}$ and $u_j \leftarrow U_j^{+} - U_j^{-}$;
17. **end for**
18. let $\pi_{t+1,j} \leftarrow \pi_{t,j-1} \times u_j$ for $j = t+1, ..., n$.
19. **end for**
20. Return the likelihood function π_{nn}.

■

Proof of Theorem 2. The characteristic function of the test statistic $SN_n(\beta_0 \mid \beta_1)$ conditional on X is given by

$$\begin{aligned}
\phi_{SN_n}(u) &= \mathbb{E}_X\left[\exp(iuSN_n(\beta_0 \mid \beta_1))\right] \\
&= \mathbb{E}_X\left[\exp\left(iu\left(\sum_{t=1}^{n} R_{t,t-1} + \sum_{t=1}^{n} \ln\left\{\frac{1-p_t[x_{t-1},\beta_0,\beta_1 \mid X]}{p_t[x_{t-1},\beta_0,\beta_1 \mid X]}\right\} s(\tilde{y}_t)\right)\right)\right],
\end{aligned}$$

which may be expressed as

$$\phi_{SN_n}(u) = \mathbb{E}_X\left[\prod_{t=1}^{n} \exp\left(iu\left(R_{t,t-1} + \ln\left\{\frac{1-p_t[x_{t-1},\beta_0,\beta_1 \mid X]}{p_t[x_{t-1},\beta_0,\beta_1 \mid X]}\right\} s(\tilde{y}_t)\right)\right)\right],$$

with $R_{1,0} = 0$, and $R_{t,t-1} = \sum_{l=t-1}^{2} \ln c_{\tilde{\sigma}_{lt}t|\tilde{\sigma}_{1t},...,\tilde{\sigma}_{t-1,t}} + \ln c_{\tilde{\sigma}_{1t}t}$ for $t = 2, ..., n$, for the D-vine-array $\tilde{A} = (\tilde{\sigma}_{lt})_{1\leq l\leq t\leq n}$, such that $l = 2, ..., n-1$ is the row with tree T_l, and column t has the permutation $\underline{\tilde{\sigma}}_{t-1} = (\tilde{\sigma}_{1t}, ..., \tilde{\sigma}_{t-1,t})$ of the previously added variables, $p_t[x_{t-1}, \beta_0, \beta_1 \mid X] = P_t[\varepsilon_t \leq f(x_{t-1}, \beta_0) - f(x_{t-1}, \beta_1) \mid X]$, and $s(\tilde{y}_t) = s(y_t - f(x_{t-1}, \beta_0))$. Furthermore, $u \in \mathbb{R}$ and the complex number $i = \sqrt{-1}$. Unlike Dufour and Taamouti (2010a), $\tilde{y}_t$ for $t = 1, ..., n$ are no longer necessarily independent conditional on X. Therefore, we follow Heinrich (1982) by expressing the characteristic function $\phi_{SN_n}(u)$ as follows

$$\phi_{SN_n}(u) = \prod_{t=1}^{n} \varphi_t(u)$$

where $\varphi_1(u) = \mathbb{E}_X \left[\exp\left(iu \left(\ln \left\{ \frac{1-p_1[x_0,\beta_0,\beta_1|X]}{p_1[x_0,\beta_0,\beta_1|X]} \right\} s(\tilde{y}_1)\right)\right)\right]$ and for $t = 2, ..., n$

$$\varphi_t(u) = \frac{f_t(u)}{f_{t-1}(u)}, \quad \text{where,} \quad f_t(u) = \mathbb{E}_X \left[\exp(iuSN_t(\beta_0 \mid \beta_1))\right]$$

Heinrich (1982) shows that $\varphi_t(u)$ can alternatively be expressed as

$$\varphi_t(u) = \mathbb{E}_X \left[\exp\left(iu \left\{R_{t,t-1} + \ln \left\{ \frac{1 - p_t[x_{t-1}, \beta_0, \beta_1 \mid X]}{p_t[x_{t-1}, \beta_0, \beta_1 \mid X]} \right\} s(\tilde{y}_t)\right\}\right)\right] + \rho_t(u)$$

where

$$\begin{aligned}
\rho_t(u) \quad = \quad & \Bigg\{\mathbb{E}_X \left[\exp\left(iu \left\{SN_t\left(\beta_0 \mid \beta_1\right)\right\}\right)\right] - \\
& \mathbb{E}_X \left[\exp\left(iu \left\{R_{t,t-1} + \ln \left\{ \frac{1 - p_t[x_{t-1}, \beta_0, \beta_1 \mid X]}{p_t[x_{t-1}, \beta_0, \beta_1 \mid X]} \right\} s(\tilde{y}_t)\right\}\right)\right] \times \\
& \mathbb{E}_X \left[\exp\left(iu \left\{SN_{t-1}\left(\beta_0 \mid \beta_1\right)\right\}\right)\right]\Bigg\} \Big/ \mathbb{E}_X \left[\exp\left(iu \left\{SN_{t-1}\left(\beta_0 \mid \beta_1\right)\right\}\right)\right].
\end{aligned}$$

Therefore, the characteristic function for the PCC-POS test statistic can be expressed as

$$\begin{aligned}
\phi_{SN_n}(u) &= \prod_{t=1}^{n} \varphi_t(u) \\
&= \prod_{t=1}^{n} \left(\mathbb{E}_X \left[\exp\left(iu \left\{R_{t,t-1} + \ln \left\{ \frac{1 - p_t[x_{t-1}, \beta_0, \beta_1 \mid X]}{p_t[x_{t-1}, \beta_0, \beta_1 \mid X]} \right\} s(\tilde{y}_t)\right\}\right)\right] + \rho_t(u)\right),
\end{aligned} \tag{2.40}$$

where $\rho_1(u) = 0$, $R_{1,0} = \rho_1(u) = 0$.

Let $Z_t = R_{t,t-1} + \ln\left\{\frac{1-p_t[x_{t-1},\beta_0,\beta_1|X]}{p_t[x_{t-1},\beta_0,\beta_1|X]}\right\} s(\tilde{y}_t)$ for $t = 1, ..., n$. Then following Heinrich (1982), and by assuming that $Z_1, ..., Z_n$ are weakly dependent, the term $\rho_t(u)$ can further be factorized. For instance, a case of such weakly dependent random variables for which a Theorem exists is the regularity Markov type process (i.e. RMT-process). Let $\mathcal{B}_s^{s+m} = \sigma(Z_s, ..., Z_{s+m})$ be the Borel σ-field generated by $\{Z_t, t = s, ..., s+u\}$. The process $\{Z_t\}_{t=1,2,...}$ is an RMT-process, if for $1 \leq s \leq t$, the uniform mixing coefficient $\phi(m) \leq \gamma(s,t)$ with probability one, where

$$\phi(m) \equiv \sup_{s\geq 1} \phi(\mathcal{B}_1^s, \mathcal{B}_{s+m}^\infty)$$

and where $\phi(\mathcal{B}_1^s, \mathcal{B}_{s+m}^\infty)$

$$\phi(\mathcal{B}_1^s, \mathcal{B}_{s+m}^\infty) \equiv \sup_{G\in\mathcal{B}_{s+m}^\infty, H\in\mathcal{B}_1^s} |P[H \mid G] - P[H]|,$$

with $\sup_{s\geq 1} \gamma(s, s+m) \to 0$ as $m \to \infty$. Given such dependence, $\rho_t(u)$ can be factorized using the results of Theorem 2 of Heinrich (1982). The conditional CDF of $SN_n(\beta_0 \mid \beta_1)$ evaluated at a constant $c_1(\beta_0, \beta_1)$, where $c_1(\beta_0, \beta_1) \in \mathbb{R}$, given by the conditional characteristic functions $\phi_{SN_n}(u)$ can then be obtained using the Fourier-inversion formula [see Gil-Pelaez (1951)] as follows

$$P[SN_n(\beta_0 \mid \beta_1) \leq c_1(\beta_0, \beta_1)] = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\mathrm{Im}\{\exp(-iuc_1(\beta_0, \beta_1))\phi_{SN_n}(u)\}}{u} du$$

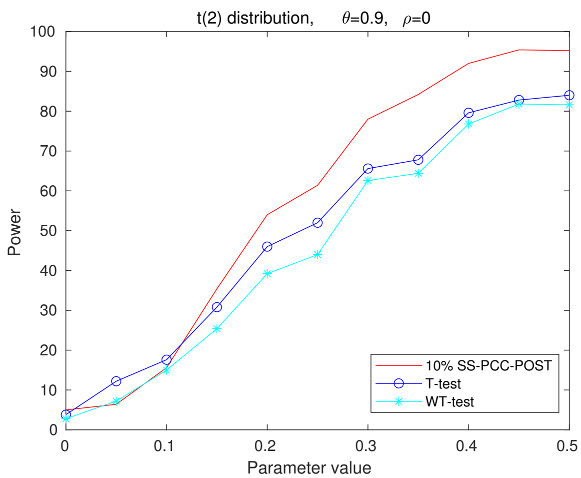
where $\forall u \in \mathbb{R}$, the conditional characteristic function $\phi_{SN_n}(u)$ is expressed by (2.40) and $\mathrm{Im}\{z\}$ denotes the imaginary part of the complex number z. Therefore, the power function can be obtained as follows

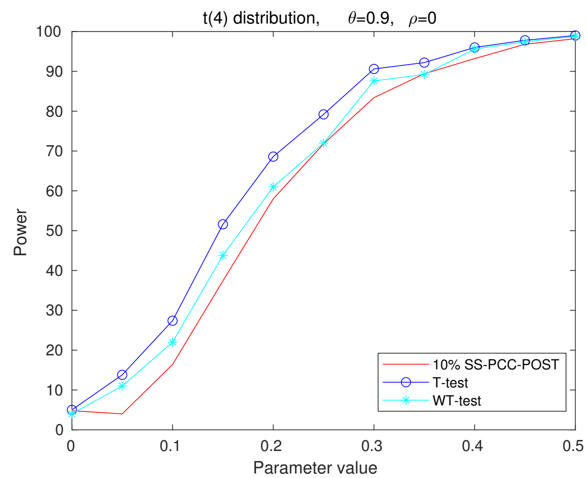
$$\Pi(\beta_0, \beta_1) = P[SN_n(\beta_0 \mid \beta_1) > c_1(\beta_0, \beta_1)] = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\mathrm{Im}\{\exp(-iuc_1(\beta_0, \beta_1))\phi_{SN_n}(u)\}}{u} du$$

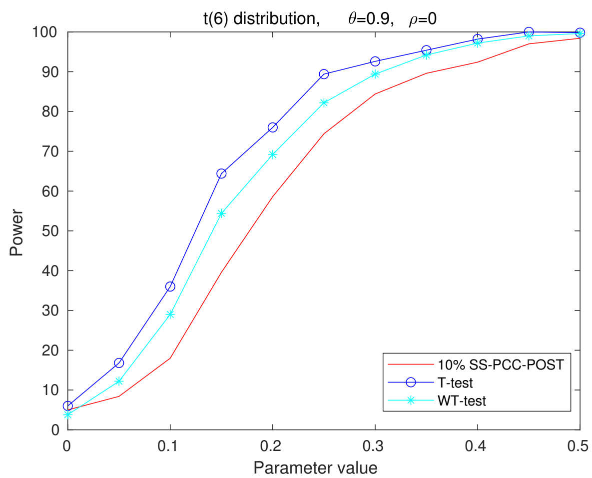
■

Additional simulations.

Figure 2.12: Power comparisons: different tests. Student's $t(\nu)$ error distributions, with different degrees of freedom ν, $\rho = 0$ in (2.33) and $\theta = 0.9$ in (2.32)

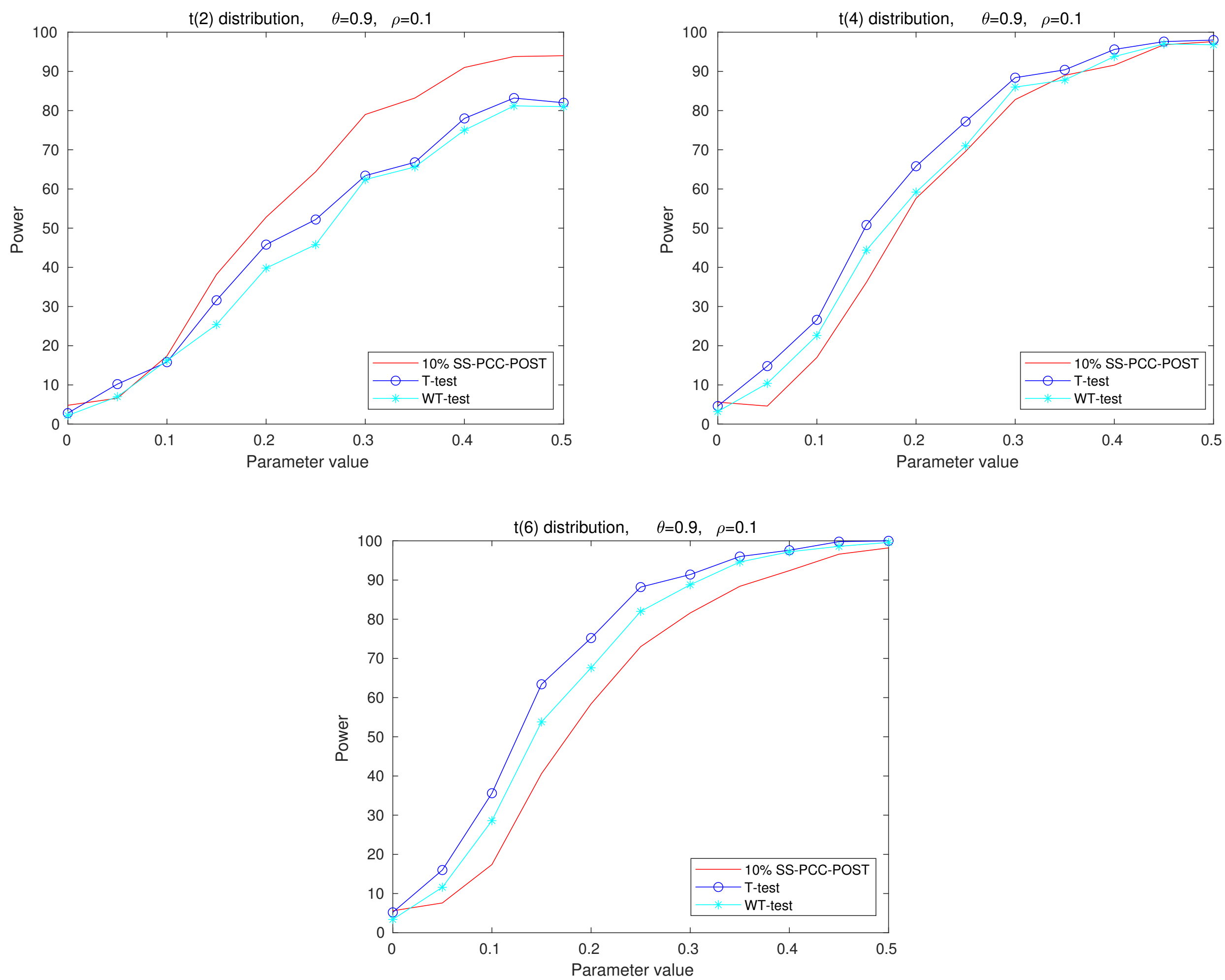






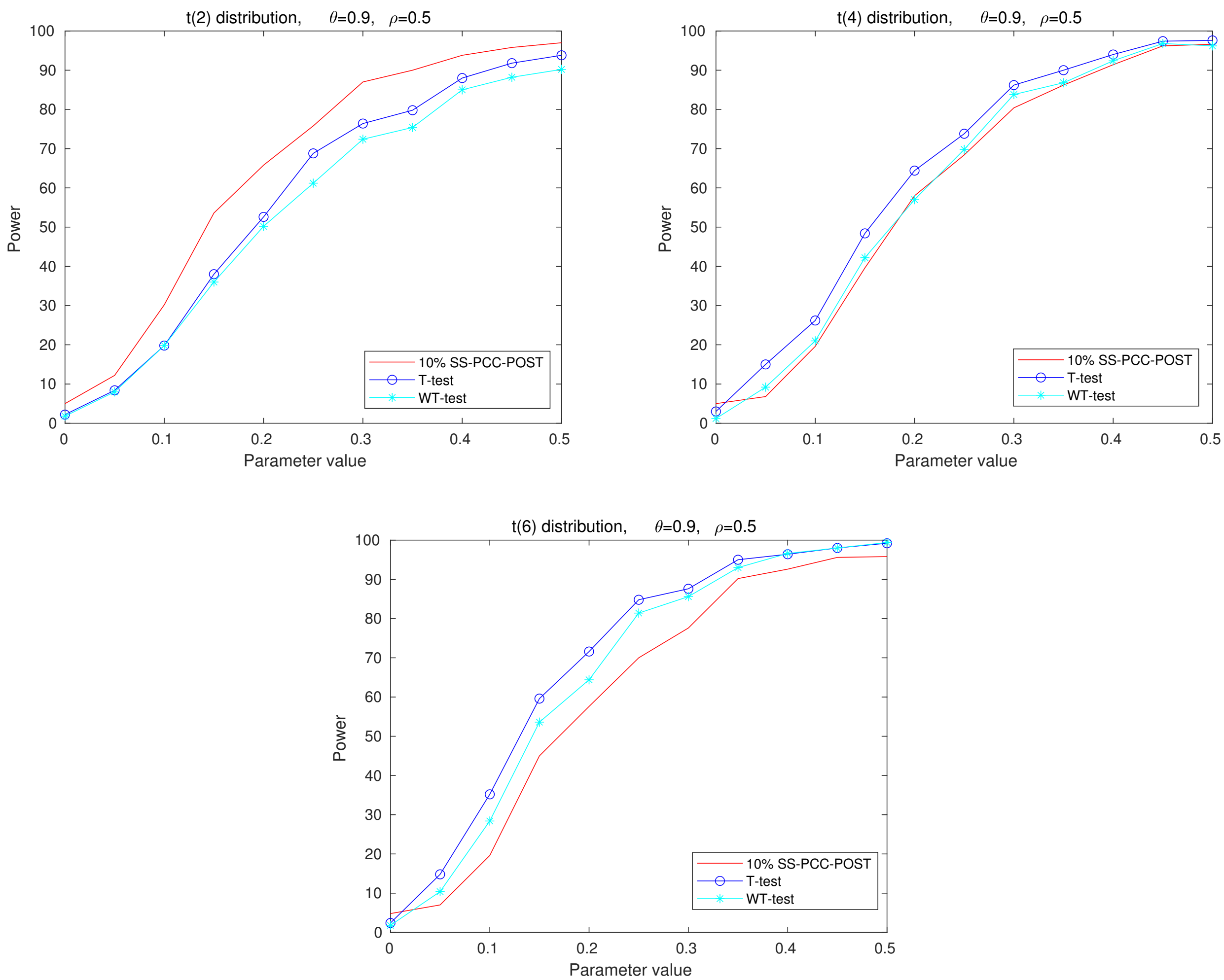
Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the T-test and (2) the T-test based on White's (1980) variance correction [WT-test].

Figure 2.13: Power comparisons: different tests. Student's $t(\nu)$ error distributions, with different degrees of freedom ν, $\rho = 0.1$ in (2.33) and $\theta = 0.9$ in (2.32)



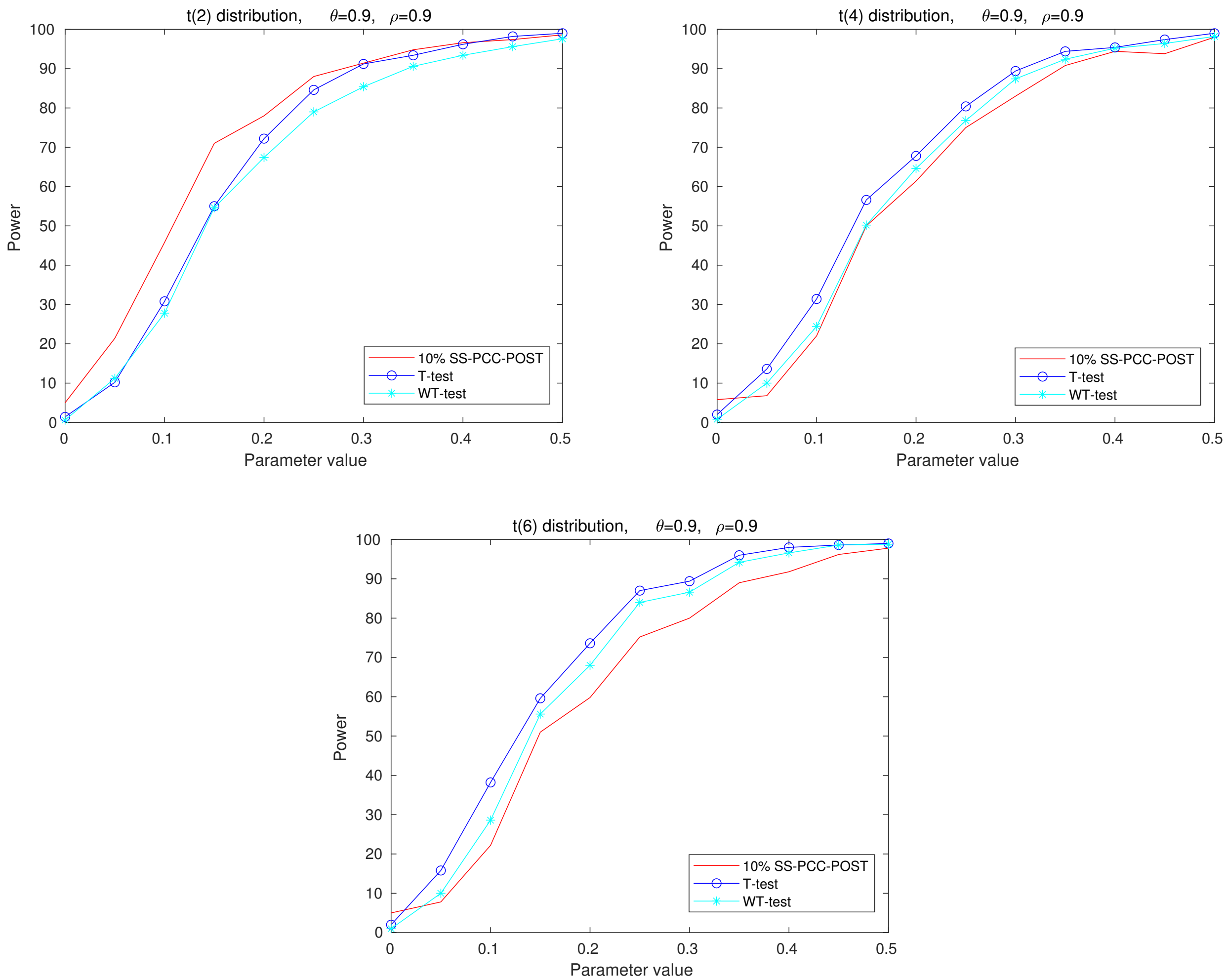
Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the T-test and (2) the T-test based on White's (1980) variance correction [WT-test].

Figure 2.14: Power comparisons: different tests. Student's $t(\nu)$ error distributions, with different degrees of freedom ν, $\rho = 0.5$ in (2.33) and $\theta = 0.9$ in (2.32)



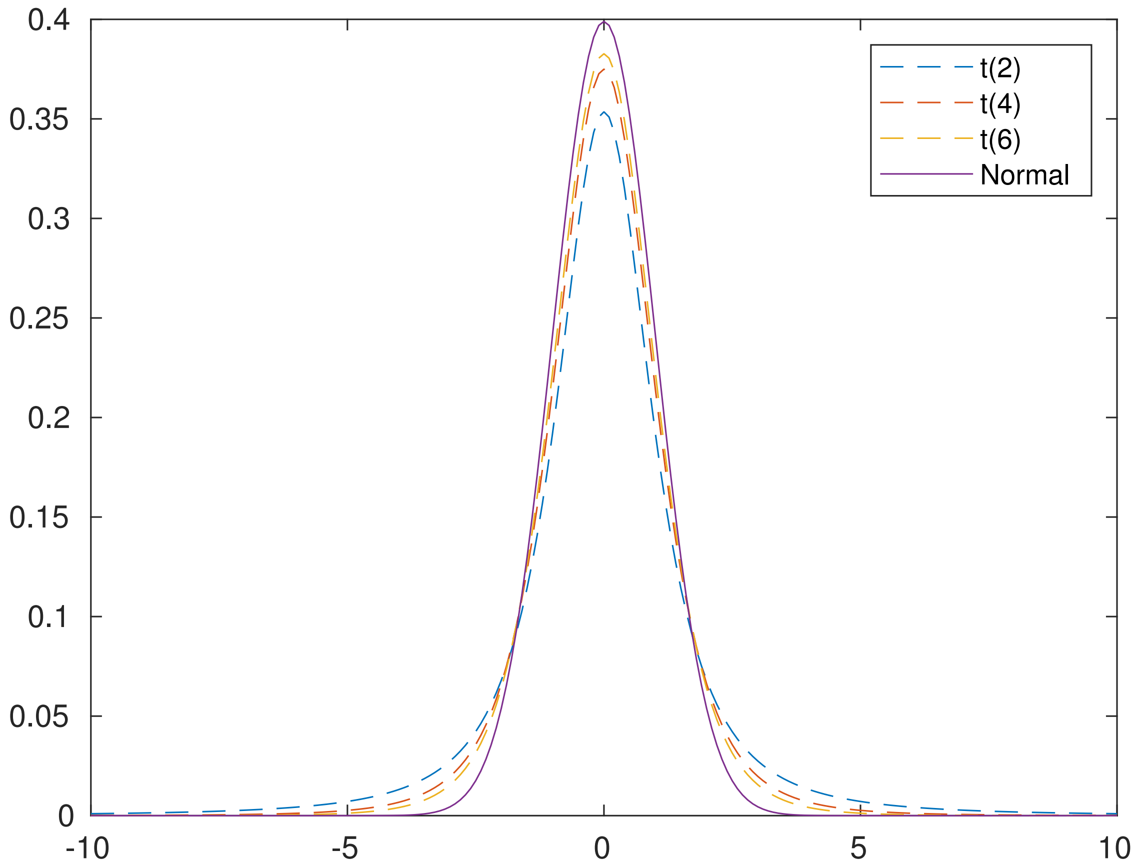
Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the T-test and (2) the T-test based on White's (1980) variance correction [WT-test].

Figure 2.15: Power comparisons: different tests. Student's $t(\nu)$ error distributions, with different degrees of freedom ν, $\rho = 0.9$ in (2.33) and $\theta = 0.9$ in (2.32)



Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the T-test and (2) the T-test based on White's (1980) variance correction [WT-test].

Figure 2.16: Comparison of the student's t distribution with various degrees of freedom to the normal distribution



Note: In this figure, we compare the Normal and Student's distribution with two, four and six degrees of freedom - i.e. $\nu = 2, \quad \nu = 4, \quad \nu = 6$.

■

Chapter 3

Sign-based Kullback measures and tests of Granger causality

3.1 Introduction

This chapter concerns the study of Wiener (1956) and Granger (1969) causality, which analyses the causal relationship between time-series. The said concept has paved the path for constructing tests and measures of Granger non-causality, where the latter has recently attracted much more attention. Earlier studies surrounding this topic concern the issue of testing Granger non-causality in parametric settings, and the measures of Granger causality initially proposed are in the context of parametric mean linear regression models [see Geweke (1982, 1984) and Dufour and Taamouti (2010b)]. Furthermore, as far as nonparametric inference is concerned, Diks and Panchenko (2006) show that the commonly used tests proposed by Hiemstra and Jones (1994) are invalid in large sample sizes and further highlight the lack of power of their own nonparametric tests against certain alternatives. Our contribution in this paper is twofold: first, we propose sign-based Granger causality measures based on the Kullback-Leibler distance to assess the strength of the causal relationships; and second, we show that by using bound-type procedures to address the nuisance parameters problem, Granger non-causality tests can be developed as a byproduct of the sign-based causality measures. These tests are exact, distribution-free and robust against heteroskedasticity of unknown form.

Wiener (1956) and Granger (1969) causality studies the predictability of, say, a (vector) variable Y from its own past and the past of another (vector) variable, say X. Granger (1969) shows that the dynamic relationship between two time-series X and Y can be broken down into three distinct types: from X to Y, from Y to X, and instantaneous causality, where all these forms of causality may coexist, reinforcing the importance of measuring the degrees of causality. Although numerous nonparametric tests of Granger non-causality have been developed to date[1], these studies only test for Granger non-causality, as opposed to measuring the degree of causality. Dufour and Taamouti (2010b) have particularly noted the study by McCloskey et al. (1996), which highlights cases where a causal effect may be large but not statistically significant, while a statistically significant effect may not have a significant impact from an economic point of view. As it has already been mentioned, measures of Granger causality have attracted much more attention in the past few decades. Measures of causality in mean have been introduced by Geweke (1982, 1984) using the mean-squared forecast errors; Polasek (1994, 2000) proposed computation of the causality measures using the Akaike and the Bayesian information criteria; Dufour and Taamouti (2010b) further expanded on the work of Geweke (1982, 1984) and introduced short and long run measures of causality in mean for vector autoregressive and moving average models; Gourieroux et al. (1987) have proposed causality measures based on the Kullback information criterion that are estimated parametrically, while Taamouti et al. (2014) use the same criterion, but estimate the measures nonparametrically using Bernstein approximation to the copula density functions to measure causality in the distribution. In a more recent study, Song and Taamouti (2018) have introduced measures of non-linear Granger causality in mean that are estimated consistently using nonparametric regression.

We introduce sign-based measures of Granger causality based on the Kullback-Leibler distance, where our measures quantify the degree of causalities. These sign-based measures are particularly attractive in cases where there is no evidence of forecastability in the mean, and a model may have better predictability power in the median rather than the upper or lower quantiles [see Furno

[1] See Hiemstra and Jones (1994), Su and White (2008), Su and White (2014), and Bouezmarni et al. (2012) for tests of causality in the distribution, and Lee and Yang (2012), Jeong et al. (2012), Hong et al. (2009) and Candelon et al. (2013) for tests of causality in quantiles.

(2014)]. However, as it has been noted by Furno (2014), in the case where the conditional distributions are symmetric, there are no significant differences between predictability in the mean and the median. As the signs exhibit serial dependence, the estimation of the sign-based measures requires the calculation of joint distribution of the signs, which is computationally infeasible. Therefore, to estimate the measures, we first impose an assumption on the sign process. Thereafter, we employ the vector autoregressive sieve bootstrap to compute the bias in finite samples and to obtain the bootstrap bias-corrected estimators of the Granger causality measures, where the validity of the vector autoregressive *sieve* boostrap is also discussed [see Meyer and Kreiss (2015) for the issue of the validity of the VAR Sieve bootstrap, and Dufour and Taamouti (2010b) and Taamouti et al. (2014) for examples of bootstrap bias-corrected estimators of causality measures]. To test the null hypothesis of Granger non-causality between random variables, we utilize the bound-type procedures as in Dufour (1990) and Campbell and Dufour (1997) to overcome the nuisance parameters hurdle and develop tests of Granger non-causality as a byproduct of the sign-based measures. The proposed tests are exact, distribution-free and robust against heteroskedasticity of unknown form. A Monte Carlo simulation study reveals that the bootstrap bias-corrected estimator of the sign-based Granger causality measures produce the desired outcome. Furthermore, the tests of Granger non-causality control size and have good power properties in finite samples. Finally, an empirical application of the measures is considered by studying the causal relationship between stock returns and the growth of the exchange rates, to illustrate the practical relevance of the sign-based causality measures and tests. Our results comprise of mixed findings; however, a bidirectional causal relationship is found between the returns of the S&P500 index and the growth of the USD/CAD exchange rates.

The structure of the paper is as follows: in Section 3.2, we outline the underlying stochastic process and define the vector of signs, as well as different hypotheses and the joint p.m.fs associated with each hypothesis. In Section 3.3, we follow Gourieroux et al. (1987) to derive the sign-based Granger causality measures. Furthermore, we present the sign-based measures by considering linear models. To simplify the calculations, we impose an assumption on the dependence structure of the sign processes. Finally, we present the estimation procedure of the sign-based measures and show the consistency of the estimators. In Section 3.4, we present the transformed model and show the bound-type testing procedure that is used for exact inference. In Section 3.5, we first evaluate the sign-based measures using long simulations. We then propose bootstrap to reduce the bias in finite samples in order to obtain a bias-corrected estimator of the measures; moreover, we discuss the asymptotic validity of the said bootstrap procedure. Finally, a Monte Carlo simulation study is conducted to show the performance of the bootstrap bias-corrected estimator of the causality measures, and the size and power properties of the proposed tests in finite samples. Section 3.6 provides an empirical application of the proposed measures. Finally, in Section 3.7 the findings of the paper are summarized.

3.2 General framework

Consider the multivariate stochastic process $Z_t = \{(Z_t^1, \cdots, Z_t^N)' : \Omega \to \mathbb{R}^N, t \geq -p+1\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. To explain the concept of Granger causality within the context of our study, we first introduce the notion of weak and strong cases of Markovian consistency [see Bielecki et al. (2008)].

Definition 2 *The stochastic process Z_t is said to satisfy the "weak"Markovian consistency of order p with respect to Z^n, if for every $B \in \mathcal{B}(\mathbb{R})$*

$$P\left[Z_t^n \in B \mid \mathcal{F}_{t-1}^{Z^n}\right] = P\left[Z_t^n \in B \mid Z_{t-1}^n, \cdots, Z_{t-p}^n\right], \quad t \geq p$$

for each $n = 1, \cdots, N$, where $\mathcal{B}(\mathbb{R})$ is a Borel set on $\mathbb{R}$.

In other words, $P[Z_t^n \leq z_t^n \mid \boldsymbol{Z}_{t-1}^n]$ depends on $\boldsymbol{Z}_{t-1}^n$ only through $Z_{t-1}^n, \cdots, Z_{t-p}^n$, for each $n = 1, \cdots, N$, where

$$\boldsymbol{Z}_{t-1}^n = \{Z_0^n, \cdots, Z_{t-1}^n\}.$$

Now let us consider the more strict case of the Markovian consistency of order p, which is defined as follows:

Definition 3 *The stochastic process Z_t is said to satisfy the "strong"Markovian consistency of* *order p with respect to Z^n, if for every $B \in \mathcal{B}(\mathbb{R})$*

$$P\left[Z_t^n \in B \mid \mathcal{F}_{t-1}^Z\right] = P\left[Z_t^n \in B \mid Z_{t-1}^n, \cdots, Z_{t-p}^n\right], \quad t \geq p$$

or equivalently

$$P\left[Z_t^n \in B \mid Z_{t-1}, \cdots, Z_0\right] = P\left[Z_t^n \in B \mid Z_{t-1}^n, \cdots, Z_{t-p}^n\right], \quad t \geq p$$

for each $n = 1, \cdots, N$, where $\mathcal{B}(\mathbb{R})$ is a Borel set on $\mathbb{R}$.

In relation to Granger causality, it can be said that when Z_t satisfies strong Markovian consistency, the collection $\{Z^i : i \neq n\}$ does not Granger cause Z^n.

For a more specific example, let $\{Z_t = (y_t, x_t)' \in \mathbb{R} \times \mathbb{R} \equiv \mathbb{R}^2, t \geq -p+1\}$ be a joint stationary stochastic process following a Markov process of order p (in the weak sense). Say, the first element of Z_t, which constitutes the process $\{x_t : t \geq -p+1\}$, has a distribution defined by the probability distribution of the initial scalar x_0, which has a density function $f(x_0)$, and the conditional probability distribution of x_t given $x_1, \cdots, x_{t-1}$, which has a conditional density function $f(x_t \mid \boldsymbol{x}_{t-1})$. Therefore. the weak case of Markovian consistency suggests that

$$f(x_t \mid \boldsymbol{x}_{t-1}) = f(x_t \mid \boldsymbol{x}_{t-1}^{t-p}), \quad t \geq p$$

whereas the strong form of Markovian consistency further suggests

$$f(x_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_{t-1}) = f(x_t \mid \boldsymbol{x}_{t-1}^{t-p}), \quad t \geq p.$$

with

$$\boldsymbol{x}_{t-1} = \{x_0, \cdots, x_{t-1}\}, \quad \text{and} \quad \boldsymbol{x}_{t-1}^{t-p} = \{x_{t-p}, \cdots, x_{t-1}\}.$$

and where $\boldsymbol{x}_{t-1}$ and $\boldsymbol{x}_{t-1}^{t-p}$ are defined identically. Hence, following Gourieroux et al. (1987) and as a consequence of the strong Markovian consistency case above, the hypotheses of Granger non-causality from X to Y and from Y to X can be expressed as

$$f(y_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_{t-1}) = f(y_t \mid \boldsymbol{y}_{t-1}), \quad t \geq p$$
$$f(x_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{x}_{t-1}) = f(x_t \mid \boldsymbol{x}_{t-1}), \quad t \geq p$$

The aim of this chapter is to derive sign-based measures and tests of Granger causality. In order to accomplish this, let the signs, $\{S_t : t \geq 1\}$, be a univariate process of binary $(0-1)$ random variables, such that

$$S_t^y = \begin{cases} 1, & y_t \geq 0 \\ 0, & y_t < 0 \end{cases}, \quad \text{and} \quad S_t^x = \begin{cases} 1, & x_t \geq 0 \\ 0, & x_t < 0 \end{cases}, \quad \text{for} \quad 1 \leq t \leq T. \tag{3.1}$$

Let the unknown true joint probability mass functions (p.m.f hereafter) of the vector of signs conditional on Y and X, under the "general"hypotheses $H^{x \to y}$ and $H^{y \to x}$ satisfy,

$$\begin{aligned} P[\mathbf{S}_T^y \mid Y, X] &= P_1[S_1^y = s_1^y \mid \mathbf{S}_0^y, Y, X] \times \prod_{t=2}^{T} P_t[S_t^y = s_t^y \mid S_{t-1}^{y,t-p}, Y, X] \\ P[\mathbf{S}_T^x \mid Y, X] &= P_1[S_1^x = s_1^x \mid \mathbf{S}_0^x, Y, X] \times \prod_{t=2}^{T} P_t[S_t^x = s_t^x \mid S_{t-1}^{x,t-p}, Y, X] \end{aligned} \tag{3.2}$$

with

$$\mathbf{S}_0^y = \{\emptyset\}, \quad \text{and} \quad P_1[S_1^y = s_1^y \mid \mathbf{S}_0^y, Y, X] = P_{0,1}[S_1^y = s_1^y \mid Y, X]$$

and where X and Y are two $(T-1) \times p$ matrices, such that

$$X = \begin{pmatrix} x_{T-1} & x_{T-2} & \cdots & x_{T-p} \\ \vdots & \vdots & \ddots & \vdots \\ x_2 & x_1 & \cdots & x_{-p+2} \\ x_1 & x_0 & \cdots & x_{-p+1} \end{pmatrix}, \quad Y = \begin{pmatrix} y_{T-1} & y_{T-2} & \cdots & y_{T-p} \\ \vdots & \vdots & \ddots & \vdots \\ y_2 & y_1 & \cdots & y_{-p+2} \\ y_1 & y_0 & \cdots & y_{-p+1} \end{pmatrix}. \tag{3.3}$$

For the rest of the paper, we mainly focus our attention on the case of causality from X to Y, noting that the procedures are identical for causality from Y to X.

Implicit in the Markovian assumption is that the p.m.f $P_t[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X]$ depends on $\mathbf{S}_{t-1}^y$ through $S_{t-1}^y, ..., S_{t-p}^y$ for all $t \geq 1$, which can alternatively be expressed as $P_t[S_t^y = s_t^y \mid S_{t-1}^{y,t-p}, Y, X]$, where $S_{t-1}^{y,t-p} = (S_{t-1}^y = s_{t-1}^y, ..., S_{t-p}^y = s_{t-p}^y)$. The joint p.m.f $P[\mathbf{S}_T^x \mid Y, X]$ belonging to $H^{y\to x}$ is symmetrical and can be expressed in a similar manner. The hypothesis of Granger non-causality from X to Y (say $H_0^{x\to y}$) implies that $\mathbf{S}_T^y$ and X are independent conditional on Y. Under this hypothesis, the joint p.m.f of $\mathbf{S}_T^y$ is expressed as

$$\begin{aligned} P[\mathbf{S}_T^y \mid Y] &= P_1[S_1^y = s_1^y \mid \mathbf{S}_0^y, Y] \times \prod_{t=2}^{T} P_t[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y] \\ &= P_1[S_1^y = s_1^y \mid \mathbf{S}_0^y, Y] \times \prod_{t=2}^{T} P_t[S_t^y = s_t^y \mid S_{t-1}^{y,t-p}, Y]. \end{aligned} \tag{3.4}$$

Therefore, the null hypothesis of Granger non-causality from X to Y, $H_0^{x\to y}$, can be expressed as

$$H_0^{x\to y} : P[\mathbf{S}_T^y \mid Y, X] = P[\mathbf{S}_T^y \mid Y]. \tag{3.5}$$

Similarly, the null hypothesis of Granger non-causality from Y to X, $H_0^{y\to x}$, can be written as

$$H_0^{y\to x} : P[\mathbf{S}_T^x \mid Y, X] = P[\mathbf{S}_T^x \mid X]. \tag{3.6}$$

The sign-based Granger causality measures define the discrepancies between the left and right hand sides of the null hypotheses (3.5) and (3.6).

3.3 Sign-based causality measures

In this Section, we derive sign-based measures of Granger causality using the Kullback-Leibler distance metric, where these derivations are inspired by Gourieroux et al. (1987). Let us assume that the unknown true probability mass function based on the signs, P, belongs to the general hypothesis $H^{x\to y}$ defined in Section 3.2 and is denoted as P_H. The Kullback-Leibler distance can be used to define the distance between the maintained and the null hypotheses - in other words, the discrepancy between the left and right hand side of (3.5). The p.m.f in H_0, by which the minimum distance is achieved is referred to as the 'pseudo-true'p.m.f. In other words, the distance between the maintained and the non-causality hypotheses can be expressed as

$$D(H/H_0) = \frac{1}{T}\min_{P_{H_0}} \mathbb{KL}(H, H_0), \tag{3.7}$$

where $\mathbb{KL}(.,.)$ denotes the Kullback-Leibler distance.

Definition 4 *Let P_H and P_{H_0} be the joint p.m.f of the signs under the maintained and the null hypotheses respectively. The Kullback-Leibler distance is defined as*

$$\mathbb{KL}(H, H_0) = \mathbb{E}_H\left\{\log\left(\frac{P_H}{P_{H_0}}\right)\right\},$$

where the expectation operator is taken on the joint p.m.f P_H.

The sign-based measure of Granger causality from X to Y, say $C(X \to Y)$, can be calculated as the difference between the discrepancies associated with the maintained hypothesis $H^{x\to y}$ and the Granger non-causality hypothesis $H_0^{x\to y}$, and with $H^{x\to y}$ itself [see Gourieroux et al. (1987)]. Formally, this may be expressed as

$$\begin{aligned} C(X \to Y) &= D(H/H_0) - D(H/H) \\ &= \frac{1}{T}\min_{P_{H_0}} \mathbb{KL}(H, H_0) - \frac{1}{T}\min_{P_H} \mathbb{KL}(H, H) \\ &= \frac{1}{T}\min_{P_{H_0}} \mathbb{KL}(H, H_0) - 0 \\ &= \frac{1}{T}\min_{P_{H_0}} \mathbb{E}_H\left\{\log\left(\frac{P_H[\mathbf{S}_T^y \mid Y, X]}{P_{H_0}[\mathbf{S}_T^y \mid Y]}\right)\right\}, \end{aligned} \tag{3.8}$$

where the sign-based measures of causality from Y to X are derived in an indentical manner. Given relationships (3.2) and (3.4), we can write the sign-based causality measure (3.8) as

$$C(X \to Y) = \tfrac{1}{T}\min_{P_{H_0}} \mathbb{E}_H\left\{\log\left[\prod_{t=1}^{T}\left(\tfrac{P_t[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X]}{P_t[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y]}\right)\right]\right\} \tag{3.9}$$

$$= \tfrac{1}{T}\min_{P_{H_0}} \mathbb{E}_H\left\{\sum_{t=1}^{T}\log\left(\tfrac{P_t[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X]}{P_t[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y]}\right)\right\}, \tag{3.10}$$

or alternatively, as

$$C(X \to Y) = \frac{1}{T} \min_{P_{H_0}} \sum_{t=1}^{T} \mathbb{E}_H \left\{ \log \left(\frac{P_t[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X]}{P_t[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y]} \right) \right\}. \tag{3.11}$$

The minimization of the Kullback-Leibler distance can be regarded as maximizing $P_{H_0}[\mathbf{S}_T^y \mid Y]$ in (3.8), or put differently as maximizing

$$\max \sum_{t=1}^{T} \mathbb{E}_H \left\{ \log \left(P_t[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y] \right) \right\} \tag{3.12}$$

in equation (3.11). The p.m.f $P_t[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y]$ is maximized only when it is equal to the 'pseudo-true' p.m.f under the null hypothesis [see Appenfix 2 of Gourieroux et al. (1987)]. As a result, the sign-based causality measure can be written as follows

$$C(X \to Y) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_H \left\{ \log \left(\frac{P_H[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X]}{P_{H_0}[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y]} \right) \right\}. \tag{3.13}$$

From the results derived above, we make the following proposition:

Proposition 4 *Assuming that under the general hypothesis, $H^{X \to Y}$, the process of signs follow a stationary process, then the Bernoulli process $\mathbf{S}_t^y$ is time invariant, and the sign-based measure of Granger causality from X to Y (3.13) can be expressed as*

$$C(X \to Y) = \mathbb{E}_H \left\{ \log \left(\frac{P_H[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X]}{P_{H_0}[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y]} \right) \right\}. \tag{3.14}$$

where similar results can be obtained for the sign-based measures of causality from Y to X.

Two important properties of the Kullback-Leibler distance that make it a desirable metric for constructing measures of Granger causality (such as measure (3.14)) are as follows:

i) The measure is non-negative; in other words

$$\mathbb{E}_H \left\{ \log \left(\frac{P_H[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X]}{P_{H_0}[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y]} \right) \right\} \geq 0$$

ii) It cancels out if and only if there is no causality.

Observe that the null hypotheses of Granger non-causality (3.5) and (3.6) can alternatively be expressed as

$$H_0^{x\rightarrow y}: P[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X] = P[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y], \quad \forall t, \tag{3.15}$$

$$H_0^{y\rightarrow x}: P[S_t^x = s_t^x \mid \mathbf{S}_{t-1}^x, Y, X] = P[S_t^x = \mathbf{s}_t^x \mid \mathbf{S}_{t-1}^x, X], \quad \forall t, \tag{3.16}$$

which implies $C(X \rightarrow Y) = 0$ or in other words that the distance between the left and right hand side of the null hypotheses (3.15) and (3.16) is zero. Therefore, large values of the measures $C(X \rightarrow Y)$ and $C(Y \rightarrow X)$ would suggest strong causality from X to Y and Y to X respectively.

3.3.1 Sign-based causality measures for linear models

Let $\{Z_t = (y_t, x_t)' \in \mathbb{R} \times \mathbb{R} \equiv \mathbb{R}^2, t \in \mathbb{Z}\}$ be a pair of covariance stationary stochastic process, such that $Z_t = (y_t, x_t)'$ is a causal linear process with a Wold representation given by

$$Z_t = \mu + \sum_{k=1}^{\infty} \psi_k \epsilon_{t-k} + \epsilon_t, \quad t \in \mathbb{Z}, \tag{3.17}$$

where $\mu \in \mathbb{R}^2$ and $\{\epsilon_t : t \in \mathbb{Z}\}$ are assumed to be bivariate and finite second moment white noise residuals, such that $\mathbb{E}(\epsilon_t) = 0$ and $\mathbb{E}(\epsilon_t \epsilon_t') = \Sigma$, where Σ is a symmetric and positive definite matrix. Under additional mild assumptions, such as the absolute summability condition of the coefficients $\psi \in \mathbb{R}^{2\times 2}$ with respect to the matrix norm $\|.\|$ (i.e. $\sum_{k=0}^{\infty} \|\psi_k\| < \infty$ with $\|\psi_k\| = \sqrt{\mathrm{tr}(\psi_k \psi_k')}$) and $\det\{\psi(z)\} \neq 0$ for all $z \in \mathbb{C}$, such that $|z| \leq 1$, with $\psi(z) = I + \sum_{k=1}^{\infty} \psi_k z^k$ and where I denotes a 2×2 idenitity matrix, it can be shown that Z_t is invertible and can be expressed as an infinite autoregressive process (i.e. VAR(∞)) as follows

$$Z_t = c + \sum_{k=1}^{\infty} \phi_k Z_{t-k} + \epsilon_t, \quad t \in \mathbb{Z}, \tag{3.18}$$

where $\sum_{k=0}^{\infty} \|\phi_k\| < \infty$ and the roots of the lag polynomial $\det\{\phi(z)\} \neq 0$ for all $z \in \mathbb{C}$, such that $|z| > 1$, and where $\phi(z) = I - \sum_{k=1}^{\infty} \phi_k z^k = \psi(z)^{-1}$. Letting $p \in \mathbb{N}$ and considering the realizations

$\{Z_0, ..., Z_T\}$, (3.18) can be approximated by a finite-order VAR(p) model, such that the order p depends on the sample size $T+1$ - i.e. $p = p(T+1)$. In other words

$$Z_t = c + \sum_{k=1}^{p} \phi_k Z_{t-k} + \epsilon_t, \quad t = p, ..., T, \tag{3.19}$$

As we are interested in measuring causality from Y to X or from X to Y, we define the marginal processes y_t and x_t which satisfy linear models of the form

$$y_t = m_1 + \sum_{k=1}^{p} a_k y_{t-k} + \sum_{k=1}^{p} b_k x_{t-k} + \varepsilon_t^y, \quad t = p, ..., T, \tag{3.20}$$

$$x_t = m_2 + \sum_{k=1}^{p} c_k y_{t-k} + \sum_{k=1}^{p} d_k x_{t-k} + \varepsilon_t^x, \quad t = p, ..., T, \tag{3.21}$$

where ε_t^y and ε_t^x are error processes satisfy the strict conditional mediangale assumption, such that

$$\varepsilon_t^y \mid Y, X \sim F(. \mid Y, X)$$

and

$$P[\varepsilon_t^y > 0 \mid \boldsymbol{\varepsilon}_{t-1}^y, Y, X] = P[\varepsilon_t^y < 0 \mid \boldsymbol{\varepsilon}_{t-1}^y, Y, X] = \frac{1}{2}, \tag{3.22}$$

with

$$\boldsymbol{\varepsilon}_0^y = \{\emptyset\}, \quad \boldsymbol{\varepsilon}_{t-1}^y = \{\varepsilon_1^y, \cdots, \varepsilon_{t-1}^y\}, \quad \text{for} \quad t \geq 2$$

where Y and X are defined similarly to the matrices (3.3), with a caveat that now the processes y_t and x_t start at $t = 0$. Furthermore, note that the strict conditional mediangale assumption is identical for process x_t. Let $\theta = (m_1, a_1, ..., a_p, b_1, ..., b_p)'$ and $\pi = (m_2, c_1, ..., c_p, d_1, ...d_p)'$ be $(2p+1) \times 1$ vectors containing the coefficients of the regression equations (3.20) and (3.21) respectively, then the regressions can be expressed as

$$y_t = \theta' J_{t-1}^p + \varepsilon_t^y, \quad t = p, ..., T, \tag{3.23}$$

$$x_t = \pi' V_{t-1}^p + \varepsilon_t^x, \quad t = p, ..., T, \tag{3.24}$$

where $J_{t-1}^{p} = (1, y_{t-1}, ..., y_{t-p}, x_{t-1}, ..., x_{t-p})'$ and $V_{t-1}^{p} = (1, y_{t-1}, ..., y_{t-p}, x_{t-1}, ..., x_{t-p})'$ are $(2p+1) \times 1$ vectors of regressors (denoted J_{t-1} and V_{t-1} hereafter). As the p.m.f of the signs now depends on the parameter vectors θ and π, it is possible to define the log-likelihood functions

$$l(U_y(T), \theta) = \sum_{t=p}^{T} \log P_\theta[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X], \tag{3.25}$$

$$l(U_x(T), \pi) = \sum_{t=p}^{T} \log P_\pi[S_t^x = s_t^x \mid \mathbf{S}_{t-1}^x, Y, X]. \tag{3.26}$$

where

$$U_y(T) = (S_p^y, ..., S_T^y)',$$

is the vector of signs for $\{y_t\}_{t=p}^{T}$, with $U_x(T)$ defined in an identical manner for $\{x_t\}_{t=p}^{T}$. In Section 3.3, we have shown that the sign-based Granger causality measure from X to Y can be regarded as the discrepancy between the maintained hypothesis $H^{x \to y}$ with true p.m.f $P_\theta[\mathbf{S}_T^y \mid Y, X]$ and the Granger non-causality hypothesis $H_0^{x \to y}$, which can formally be expressed as

$$C(X \to Y) = \frac{1}{T-p+1} \min_{\theta \in H_0^{x \to y}} \mathbb{E}_H \left\{ \log \left(\frac{P_\theta[\mathbf{S}_T^y \mid Y, X]}{P_{\theta^R}[\mathbf{S}_T^y \mid Y]} \right) \right\}, \tag{3.27}$$

where θ^R is the pseudo-true value of θ and where the measure of causality from Y to X is expressed conversely.

3.3.2 Estimation

Let us denote the unconstrained and constrained finite-sample OLS estimates of θ, by $\hat{\theta}$ and $\hat{\theta}^R$ respectively, such that $\hat{\theta} = (\hat{m}_1, \hat{a}_1, ..., \hat{a}_p, \hat{b}_1, ..., \hat{b}_p)'$ and $\hat{\theta}^R = (\hat{m}_1^R, \hat{a}_1^R, ..., \hat{a}_p^R, \underbrace{0, ..., 0}_{p})'$ is the OLS estimate of the restricted model

$$y_t = m_1^R + \sum_{k=1}^{p} a_k^R y_{t-k} + \varepsilon_t^y, \quad t = p, ..., T, \tag{3.28}$$

respectively. For convenience Dufour and Taamouti (2010b) recommend estimating the constrained and the unrestricted models with the same lag order p. Then a natural estimator for

$C(X \to Y)$ is

$$\hat{C}(X \to Y) = \frac{1}{T-p+1}\left[l(U_y(T), \hat{\theta}) - l(U_y(T), \hat{\theta}^R)\right]. \quad (3.29)$$

From (3.25) it is clear that the estimator of the sign-based measure of causality from X to Y (3.29) can alternatively be expressed as

$$\hat{C}(X \to Y) = \frac{1}{T-p+1}\sum_{t=p}^{T}\left\{\log P_{\hat{\theta}}[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X] - \log P_{\hat{\theta}^R}[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y]\right\}, \quad (3.30)$$

where as in the first chapter

$$\log P_{\hat{\theta}}[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X] = S_t^y \log\left\{\frac{P_{\hat{\theta}}[y_t \geq 0 \mid \mathbf{S}_{t-1}^y, Y, X]}{P_{\hat{\theta}}[y_t < 0 \mid \mathbf{S}_{t-1}^y, Y, X]}\right\} + \log P_{\hat{\theta}}[y_t < 0 \mid \mathbf{S}_{t-1}^y, Y, X], \quad (3.31)$$

and such that $\log P_{\hat{\theta}^R}[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X]$ for the constrained regression is expressed identically. As before, the estimator of the sign-based measures requires the calculation of the joint distribution of the signs, which is computationally infeasible. Therefore, we impose the following assumption:

Assumption A3: *Let $\{y_t, t = p, p+1, \cdots\}$ follow a Markov process of order one. Then the signs $\{S_t\}_{t=p}^{\infty}$ follow a Markov process of the same order and the p.m.fs associated with the unrestricted and constrained regressions can be expressed as*

$$\begin{cases} P[y_t \geq 0 \mid \mathbf{S}_{t-1}, .] = P[y_t \geq 0 \mid y_{t-1} \geq 0, .]^{S_{t-1}} P[y_t \geq 0 \mid y_{t-1} < 0, .]^{1-S_{t-1}} \\ P[y_t < 0 \mid \mathbf{S}_{t-1}, .] = P[y_t < 0 \mid y_{t-1} \geq 0, .]^{S_{t-1}} P[y_t < 0 \mid y_{t-1} < 0, .]^{1-S_{t-1}} \end{cases}$$

Given this assumption, we introduce the following corollary.

Corollary 5 *From proposition 4 and given assumption A3, it follows that the estimator of the sign-based Granger causality measure from X to Y under the assumptions (3.24) and (3.22) can be calculated by*

$$\hat{C}(X \to Y) = \frac{1}{T-p+1}\sum_{t=p}^{T}\left\{S_t^y S_{t-1}^y \alpha_t(\hat{\theta}/\hat{\theta}^R) + S_t^y \beta_t(\hat{\theta}/\hat{\theta}^R) + S_{t-1}^y \gamma_t(\hat{\theta}/\hat{\theta}^R) + \delta_t(\hat{\theta}/\hat{\theta}^R)\right\}, \quad (3.32)$$

where

$$\alpha_p(\hat{\theta}/\hat{\theta}^R) = 0, \quad \gamma_p(\hat{\theta}/\hat{\theta}^R) = 0$$

and

$$\begin{aligned}
\beta_p(\hat{\theta}/\hat{\theta}^R) &= \log\left\{\frac{(1 - P[\varepsilon_p < -\hat{\theta}' J_0 \mid Y, X])P[\varepsilon_p < -\hat{\theta}'^R J_0 \mid Y]}{(1 - P[\varepsilon_p < -\hat{\theta}'^R J_0 \mid Y])P[\varepsilon_p < -\hat{\theta}' J_0 \mid Y, X]}\right\} \\
\delta_p(\hat{\theta}/\hat{\theta}^R) &= \log\left\{\frac{P[\varepsilon_p < -\hat{\theta}' J_0 \mid Y, X]}{P[\varepsilon_p < -\hat{\theta}'^R J_0 \mid Y]}\right\}
\end{aligned}$$

and where for $t = p + 1, ..., T$

$$\begin{aligned}
\alpha_t(\hat{\theta}/\hat{\theta}^R) &= \Bigg[\Bigg(\log\left\{\frac{1 - \left(\frac{P[\varepsilon_t < -\hat{\theta}' J_{t-1}|Y,X]}{1 - P[\varepsilon_{t-1} < -\hat{\theta}' J_{t-2}|Y,X]} - \frac{P[\varepsilon_t < -\hat{\theta}' J_{t-1}, \varepsilon_{t-1} < -\hat{\theta}' J_{t-2}|Y,X]}{1 - P[\varepsilon_{t-1} < -\hat{\theta}' J_{t-2}|Y,X]}\right)}{\frac{P[\varepsilon_t < -\hat{\theta}' J_{t-1}|Y,X]}{1 - P[\varepsilon_{t-1} < -\hat{\theta}' J_{t-2}|Y,X]} - \frac{P[\varepsilon_{t-1} < -\hat{\theta}' J_{t-2}, \varepsilon_t < -\hat{\theta}' J_{t-1}|Y,X]}{1 - P[\varepsilon_{t-1} < -\hat{\theta}' J_{t-2}|Y,X]}}\right\} \\
&\quad - \log\left\{\frac{1 - \frac{P[\varepsilon_t < -\hat{\theta}' J_{t-1}, \varepsilon_{t-1} < -\hat{\theta}' J_{t-2}|Y,X]}{P[\varepsilon_{t-1} < -\hat{\theta}' J_{t-2}|Y,X]}}{\frac{P[\varepsilon_t < -\hat{\theta}' J_{t-1}, \varepsilon_{t-1} < -\hat{\theta}' J_{t-2}|Y,X]}{P[\varepsilon_{t-1} < -\hat{\theta}' J_{t-2}|Y,X]}}\right\}\Bigg) \\
&\quad - \Bigg(\log\left\{\frac{1 - \left(\frac{P[\varepsilon_t < -\hat{\theta}'^R J_{t-1}|Y]}{1 - P[\varepsilon_{t-1} < -\hat{\theta}'^R J_{t-2}|Y]} - \frac{P[\varepsilon_{t-1} < -\hat{\theta}'^R J_{t-2}, \varepsilon_t < -\hat{\theta}'^R J_{t-1}|Y]}{1 - P[\varepsilon_{t-1} < -\hat{\theta}'^R J_{t-2}|Y]}\right)}{\frac{P[\varepsilon_t < -\hat{\theta}'^R J_{t-1}|Y]}{1 - P[\varepsilon_{t-1} < -\hat{\theta}'^R J_{t-2}|Y]} - \frac{P[\varepsilon_{t-1} < -\hat{\theta}'^R J_{t-2}, \varepsilon_t < -\hat{\theta}'^R J_{t-1}|Y]}{1 - P[\varepsilon_{t-1} < -\hat{\theta}'^R J_{t-2}|Y]}}\right\} \\
&\quad - \log\left\{\frac{1 - \frac{P[\varepsilon_t < -\hat{\theta}'^R J_{t-1}, \varepsilon_{t-1} < -\hat{\theta}'^R J_{t-2}|Y]}{P[\varepsilon_{t-1} < -\hat{\theta}'^R J_{t-2}|Y]}}{\frac{P[\varepsilon_t < -\hat{\theta}'^R J_{t-1}, \varepsilon_{t-1} < -\hat{\theta}'^R J_{t-2}|Y]}{P[\varepsilon_{t-1} < -\hat{\theta}'^R J_{t-2}|Y]}}\right\}\Bigg)\Bigg] \\
\beta_t(\hat{\theta}/\hat{\theta}^R) &= \left[\log\left\{\frac{1 - \frac{P[\varepsilon_t < -\hat{\theta}' J_{t-1}, \varepsilon_{t-1} < -\hat{\theta}' J_{t-2}|Y,X]}{P[\varepsilon_{t-1} < -\hat{\theta}' J_{t-2}|Y,X]}}{\frac{P[\varepsilon_t < -\hat{\theta}' J_{t-1}, \varepsilon_{t-1} < -\hat{\theta}' J_{t-2}|Y,X]}{P[\varepsilon_{t-1} < -\hat{\theta}' J_{t-2}|Y,X]}}\right\}\right. \\
&\quad \left. - \log\left\{\frac{1 - \frac{P[\varepsilon_t < -\hat{\theta}'^R J_{t-1}, \varepsilon_{t-1} < -\hat{\theta}'^R J_{t-2}|Y]}{P[\varepsilon_{t-1} < -\hat{\theta}'^R J_{t-2}|Y]}}{\frac{P[\varepsilon_t < -\hat{\theta}'^R J_{t-1}, \varepsilon_{t-1} < -\hat{\theta}'^R J_{t-2}|Y]}{P[\varepsilon_{t-1} < -\hat{\theta}'^R J_{t-2}|Y]}}\right\}\right]
\end{aligned}$$

$$
\begin{aligned}
\gamma_t(\hat{\theta}/\hat{\theta}^R) &= \Bigg[\log\left\{\frac{\frac{P[\varepsilon_t<-\hat{\theta}'J_{t-1}|Y,X]}{1-P[\varepsilon_{t-1}<-\hat{\theta}'J_{t-2}|Y,X]}-\frac{P[\varepsilon_{t-1}<-\hat{\theta}'J_{t-2},\varepsilon_t<-\hat{\theta}'J_{t-1}|Y,X]}{1-P[\varepsilon_{t-1}<-\hat{\theta}'J_{t-2}|Y,X]}}{\frac{P[\varepsilon_t<-\hat{\theta}'J_{t-1},\varepsilon_{t-1}<-\hat{\theta}'J_{t-2}|Y,X]}{P[\varepsilon_{t-1}<-\hat{\theta}'J_{t-2}|Y,X]}}\right\} \\
&\quad -\log\left\{\frac{\frac{P[\varepsilon_t<-\hat{\theta}'^RJ_{t-1}|Y]}{1-P[\varepsilon_{t-1}<-\hat{\theta}'^RJ_{t-2}|Y]}-\frac{P[\varepsilon_{t-1}<-\hat{\theta}'^RJ_{t-2},\varepsilon_t<-\hat{\theta}'^RJ_{t-1}|Y]}{1-P[\varepsilon_{t-1}<-\hat{\theta}'^RJ_{t-2}|Y]}}{\frac{P[\varepsilon_t<-\hat{\theta}'^RJ_{t-1},\varepsilon_{t-1}<-\hat{\theta}'^RJ_{t-2}|Y]}{P[\varepsilon_{t-1}<-\hat{\theta}'^RJ_{t-2}|Y]}}\right\}\Bigg] \\
\delta_t(\hat{\theta}/\hat{\theta}^R) &= \log\left\{\frac{\frac{P[\varepsilon_t<-\hat{\theta}'J_{t-1},\varepsilon_{t-1}<-\hat{\theta}'J_{t-2}|Y,X]}{P[\varepsilon_{t-1}<-\hat{\theta}'J_{t-2}|Y,X]}}{\frac{P[\varepsilon_t<-\hat{\theta}'^RJ_{t-1},\varepsilon_{t-1}<-\hat{\theta}'^RJ_{t-2}|Y]}{P[\varepsilon_{t-1}<-\hat{\theta}'^RJ_{t-2}|Y]}}\right\}
\end{aligned}
$$

Proof: *See Appendix.*

A special case is where $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{T-1}, \varepsilon_T$ are distributed according to $N(0,1)$. As suggested before, since the form of the serial dependence of the errors is non-linear, we may calculate the bivariate probabilities using "jointly-symmetric"copulae introduced in Chapter 1, by considering the Archimedean Frank, Clayton or Gumbel as the copula family [see Joe (2014)]. Alternatively, we may evaluate the bivariate probabilities $P[\ \varepsilon_{t-1} < \cdot,\ \varepsilon_t < \cdot \mid X]$ using a multivariate Student's t distribution by imposing the identity matrix I. Then the estimated weights $\alpha_t(\hat{\theta}/\hat{\theta}^R)$, $\beta_t(\hat{\theta}/\hat{\theta}^R)$, $\gamma_t(\hat{\theta}/\hat{\theta}^R)$, and $\delta_t(\hat{\theta}/\hat{\theta}^R)$ for the sign-based Granger causality measure from X to Y (3.32) are given by

$$
\alpha_p(\hat{\theta}/\hat{\theta}^R) = 0, \quad \gamma_p(\hat{\theta}/\hat{\theta}^R) = 0
$$

and

$$
\beta_p(\hat{\theta}/\hat{\theta}^R) = \log\left\{\frac{\Phi(\hat{\theta}'J_0)(1-\Phi(\hat{\theta}'^RJ_0))}{\Phi(\hat{\theta}'^RJ_0)(1-\Phi(\hat{\theta}'J_0))}\right\}, \quad \delta_p(\hat{\theta}/\hat{\theta}^R) = \log\left\{\frac{1-\Phi(\hat{\theta}'J_0)}{1-\Phi(\hat{\theta}'^RJ_0)}\right\}
$$

and where for $t = p+1, ..., T$

$$
\begin{aligned}
\alpha_t(\hat{\theta}/\hat{\theta}^R) &= \Bigg[\Bigg(\log\left\{\frac{1-\left(\frac{1-\Phi(\hat{\theta}'J_{t-1})}{\Phi(\hat{\theta}'J_{t-2})}-\frac{C^{JS}(\Phi(-\hat{\theta}'J_{t-1}),\Phi(-\hat{\theta}'J_{t-2}))}{\Phi(\hat{\theta}'J_{t-2})}\right)}{\frac{1-\Phi(\hat{\theta}'J_{t-1})}{\Phi(\hat{\theta}'J_{t-2})}-\frac{C^{JS}(\Phi(-\hat{\theta}'J_{t-2}),\Phi(-\hat{\theta}'J_{t-1}))}{\Phi(\hat{\theta}'J_{t-2})}}\right\}-\log\left\{\frac{1-\frac{C^{JS}(\Phi(-\hat{\theta}'J_{t-1}),\Phi(-\hat{\theta}'J_{t-2}))}{1-\Phi(\hat{\theta}'J_{t-2})}}{\frac{C^{JS}(\Phi(-\hat{\theta}'J_{t-1}),\Phi(-\hat{\theta}'J_{t-2}))}{1-\Phi(\hat{\theta}'J_{t-2})}}\right\}\Bigg) \\
&\quad -\Bigg(\log\left\{\frac{1-\left(\frac{1-\Phi(\hat{\theta}'^RJ_{t-1})}{\Phi(\hat{\theta}'^RJ_{t-2})}-\frac{C^{JS}(\Phi(-\hat{\theta}'^RJ_{t-2}),\Phi(-\hat{\theta}'^RJ_{t-1}))}{\Phi(\hat{\theta}'^RJ_{t-2})}\right)}{\frac{1-\Phi(\hat{\theta}'^RJ_{t-1})}{\Phi(\hat{\theta}'^RJ_{t-2})}-\frac{C^{JS}(\Phi(-\hat{\theta}'^RJ_{t-2}),\Phi(-\hat{\theta}'^RJ_{t-1}))}{\Phi(\hat{\theta}'^RJ_{t-2})}}\right\} \\
&\quad -\log\left\{\frac{1-\frac{C^{JS}(\Phi(-\hat{\theta}'^RJ_{t-1}),\Phi(-\hat{\theta}'^RJ_{t-2}))}{1-\Phi(\hat{\theta}'^RJ_{t-2})}}{\frac{C^{JS}(\Phi(-\hat{\theta}'^RJ_{t-1}),\Phi(-\hat{\theta}'^RJ_{t-2}))}{1-\Phi(\hat{\theta}'^RJ_{t-2})}}\right\}\Bigg)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\beta_t(\hat{\theta}/\hat{\theta}^R) &= \left[\log\left\{\frac{1-\frac{C^{JS}(\Phi(-\hat{\theta}'J_{t-1}),\Phi(-\hat{\theta}'J_{t-2}))}{1-\Phi(\hat{\theta}'J_{t-2})}}{\frac{C^{JS}(\Phi(-\hat{\theta}'J_{t-1}),\Phi(-\hat{\theta}'J_{t-2}))}{1-\Phi(\hat{\theta}'J_{t-2})}}\right\} - \log\left\{\frac{1-\frac{C^{JS}(\Phi(-\hat{\theta}'^RJ_{t-1}),\Phi(-\hat{\theta}'^RJ_{t-2}))}{1-\Phi(\hat{\theta}'^RJ_{t-2})}}{\frac{C^{JS}(\Phi(-\hat{\theta}'^RJ_{t-1}),\Phi(-\hat{\theta}'^RJ_{t-2}))}{1-\Phi(\hat{\theta}'^RJ_{t-2})}}\right\}\right] \\
\gamma_t(\hat{\theta}/\hat{\theta}^R) &= \left[\log\left\{\frac{\frac{1-\Phi(\hat{\theta}'J_{t-1})}{\Phi(\hat{\theta}'J_{t-2})}-\frac{C^{JS}(\Phi(-\hat{\theta}'J_{t-2}),\Phi(-\hat{\theta}'J_{t-1}))}{\Phi(\hat{\theta}'J_{t-2})}}{\frac{C^{JS}(\Phi((-\hat{\theta}'J_{t-1}),\Phi(-\hat{\theta}'J_{t-2}))}{1-\Phi(\hat{\theta}'J_{t-2})}}\right\}\right. \\
&\quad \left. -\log\left\{\frac{\frac{1-\Phi(\hat{\theta}'^RJ_{t-1})}{\Phi(\hat{\theta}'^RJ_{t-2})}-\frac{C^{JS}(\Phi(-\hat{\theta}'^RJ_{t-2}),\Phi(-\hat{\theta}'^RJ_{t-1}))}{\Phi(\hat{\theta}'^RJ_{t-2})}}{\frac{C^{JS}(\Phi(-\hat{\theta}'^RJ_{t-1}),\Phi(-\hat{\theta}'^RJ_{t-2}))}{1-\Phi(\hat{\theta}'^RJ_{t-2})}}\right\}\right] \\
\delta_t(\hat{\theta}/\hat{\theta}^R) &= \log\left\{\frac{\frac{C^{JS}(\Phi(-\hat{\theta}'J_{t-1}),\Phi(-\hat{\theta}'J_{t-2}))}{1-\Phi(\hat{\theta}'J_{t-2})}}{\frac{C^{JS}(\Phi(-\hat{\theta}'^RJ_{t-1}),\Phi(-\hat{\theta}'^RJ_{t-2}))}{1-\Phi(\hat{\theta}'^RJ_{t-2})}}\right\}
\end{aligned}
$$

where $\Phi(.)$ is the standard normal distribution function and $C^{JS}(u_1, u_2)$ is the "jointly-symmetric" bivariate copula with uniformly distributed marginals. To prove the consistency of the above estimator some regularity conditions are needed. We present a set of assumptions considered by White (2014) and Andrews (1992) among others, which has been adopted by Coudin and Dufour (2004) in the context of sign-based estimators. The following conditions are satisfied:

Assumption A4:

(1) **Mixing.** Let $\{(J'_t, \varepsilon^y_t) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R} \equiv \mathbb{R}^3, t = 0, 1, 2, ...\}$. *Then* $\{(J'_t, \varepsilon^y_t)\}$ *is* α*-mixing of size* $-r/(r-1)$ *with* $r > 1$.

(2) **Exogeneity.** $\mathbb{E}(J_{t-1}\varepsilon^y_t) = 0 \quad t = 1, 2, ...$

(3) **Boundedness.** i) $\mathbb{E} \mid x_t \mid^{r+\delta} < \Delta < \infty$ *for some* $\delta > 0$ *and* $\forall t \in \mathbb{N}$ ii) $\mathbb{E} \mid y_t \mid^{r+\delta} < \Delta < \infty$ *for some* $\delta > 0$ *and* $\forall t \in \mathbb{N}$ iii) $\mathbb{E} \mid x_{t-1}\varepsilon^y_t \mid^{r+\delta} < \Delta < \infty$ *for some* $\delta > 0$ *and* $\forall t \in \mathbb{N}$ iv) $\mathbb{E} \mid y_{t-1}\varepsilon^y_t \mid^{r+\delta} < \Delta < \infty$ *for some* $\delta > 0$ *and* $\forall t \in \mathbb{N}$ v) $\mathbb{E} \mid x^2_t \mid^{r+\delta} < \Delta < \infty$ *for some* $\delta > 0$ *and* $\forall t \in \mathbb{N}$. vi) $\mathbb{E} \mid y^2_t \mid^{r+\delta} < \Delta < \infty$ *for some* $\delta > 0$ *and* $\forall t \in \mathbb{N}$.

(4) **Positive definiteness.** *Let* $\mathbf{J} = (y_{T-1}, x_{T-1}, ..., y_p, x_p)' \in \mathbb{R}^{(T-p+1)\times 2p}$. *Then* $M_n \equiv \mathbb{E}(\mathbf{J}'\mathbf{J}/n)$ *is uniformly positive definite, where* $n = T - p + 1$.

(5) **Compactness.** *Parameter space* Θ *is compact.*

Proposition 5 *Under assumptions A4, the estimator* $\hat{C}(X \to Y)$ *converges in probability to* $C(X \to Y)$.

Proof: *See Appendix.*

3.4 Inference

Exact tests of Granger non-causality between random variables can be developed as a byproduct of the sign-based measures proposed in Section 3.3.2. Many exact parametric tests assume that the residuals follow a Gaussian distribution, which is unrealistic in the presence of data with heavy tailed and asymmetric distributions. As a result, these tests may not perform well in terms of size control and power [see. Dufour and Taamouti (2010a)].

We propose exact and distribution-free tests of Granger non-causality that are robust against heteroskedasticity of unknown form. We first consider the testing problem with known nuisance parameters. Then the results are extended to scenarios where the nuisance parameters are unknown. To address the latter, we adopt the bound-type procedure as in Dufour (1990) and Campbell and Dufour (1997) to remedy the nuisance parameter problem under the null hypothesis of Granger non-causality.

Once again we focus on the causality from X to Y and consider the regression equation

$$y_t = \theta' J_{t-1}^p + \varepsilon_t^y, \quad t = p, ..., T, \tag{3.33}$$

where as before $\theta = (m_1, a_1, ..., a_p, b_1, ..., b_p)'$ and $J_{t-1} = (1, y_{t-1}, ..., y_{t-p}, x_{t-1}, ..., x_{t-p})'$. Assume we wish to test the null hypothesis

$$H_0^{x \to y} : C(X \to Y) = 0, \tag{3.34}$$

We know that under the null hypothesis of Granger non-causality the regression model is constrained with $\theta^R = (m_1^R, a_1^R, ..., a_p^R, \underbrace{0, ..., 0}_{p})'$. Therefore, testing the null hypothesis (3.34) in the context of regression (3.33) is equivalent to testing

$$H_0 : \theta = \theta^R. \tag{3.35}$$

Notice that the estimator (3.29) of the sign-based measures based on log-likelihood functions for the regression equation (3.33) can be expressed as

$$\hat{C}(X \to Y) = \frac{1}{T-p+1} \left\{ l(U_y(n), \hat{\theta}) - l(U_y(n), \hat{\theta}^R) \right\},$$

which is of similar form to the Neyman-Pearson type tests based on the signs introduced in the previous chapters. Therefore, a test based on the signs for testing the Granger non-causality hypothesis (3.35) against an alternative

$$H_1 : \theta = \theta_1, \quad \theta_1 \neq 0, \tag{3.36}$$

can be constructed in a similar manner to the previous chapters. However, we provide a caveat that the test statistic that corresponds to this testing problem allows for unknown nuisance parameters θ^R under the null hypothesis of Granger non-causality; hence, unlike the preceding chapters the test statistic is not a *pivotal function*, as it depends on the distribution of the residuals ε_t. On the other hand, if the model is transformed such that there are no nuisance parameters under the null hypothesis, then the test statistic is distribution-free. Therefore, we propose a transformation such that the test is distribution-free and the dependent variable has zero median under the null. In what follows, we first consider making inference in the case where the vector of nuisance parameters θ^R is known. We then proceed to show how by using Bonferroni-type tests, it is possible to make provably valid inference with unknown nuisance parameters.

3.4.1 Inference with *known* nuisance parameters

Let us rewrite regression equation (3.33) as

$$y_t = A' \underline{y}_{t-1} + \beta' \underline{x}_{t-1} + \varepsilon_t, \quad t = p, ..., T, \tag{3.37}$$

such that $\underline{y}_{t-1} = (1, y_{t-1}, ..., y_{t-p})'$ and $\underline{x}_{t-1} = (x_{t-1}, ..., x_{t-p})$, are in turn $(p+1) \times 1$ and $p \times 1$ vector of regressors, and $A = (m_1, a_1, ..., a_p)'$ and $\beta = (b_1, ..., b_p)'$ are $(p+1) \times 1$ and $p \times 1$ vector of unknown parameters. Finally, ε_t for $p \leq t \leq T$ are the error terms with conditional distribution function

$$\varepsilon_t \mid Y, X \sim F_t(. \mid Y, X), \tag{3.38}$$

where $F_t(. \mid Y, X)$ is a distribution function, where the process ε_t satisfies the strict conditional mediangale assumption

$$P[\varepsilon_t > 0 \mid \boldsymbol{\varepsilon}_{t-1}, Y, X] = P[\varepsilon_t < 0 \mid \boldsymbol{\varepsilon}_{t-1}, Y, X] = \frac{1}{2}. \tag{3.39}$$

where

$$\boldsymbol{\varepsilon}_0 = \{\emptyset\}, \quad \boldsymbol{\varepsilon}_{t-1} = \{\varepsilon_1, \cdots, \varepsilon_{t-1}\}$$

Notice that regression (3.37) can be transformed as follows

$$\tilde{y}_t = \beta x_{t-1} + \varepsilon_t, \quad t = p, ..., T, \tag{3.40}$$

where

$$\tilde{y}_t = y_t - A\underline{y}_{t-1}, \quad t = p, ..., T. \tag{3.41}$$

such that A is taken to be $A^R = (m_1^R, a_1^R, ..., a_p^R)$ when the coefficient A^R is known under the restricted model. Based on the transformation, we now define a new vector of signs

$$S_t^{\tilde{y}} = \begin{cases} 1 \ \ if \ \ \tilde{y}_t \geq 0 \\ 0 \ \ if \ \ \tilde{y}_t < 0 \end{cases}, \quad t = p, ..., T. \tag{3.42}$$

Therefore, testing the null hypothesis of Granger non-causality (3.35) is equivalent to testing the null hypothesis

$$\tilde{H}_0^{x \to y} : \beta = 0, \tag{3.43}$$

against the alternative

$$\tilde{H}_1^{x \to y} : \beta = \beta_1, \quad \beta_1 \neq 0. \tag{3.44}$$

Under the assumption (3.39), and given the transformation (3.40), the estimator of the sign-based measure from X to Y for the regression equation (3.37) can now be written as

$$\begin{aligned}\hat{C}(X \to Y) &= \frac{1}{T-p+1}\left[l(U_{\tilde{y}}(T), \beta_1) - l(U_{\tilde{y}}(T), 0)\right] \\ &= \frac{1}{T-p+1}\sum_{t=p}^{T}\left\{\log P_{\beta_1}[S_t^{\tilde{y}} = s_t^{\tilde{y}} \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X] - \log P_0[S_t^{\tilde{y}} = s_t^{y} \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y]\right\},\end{aligned} \tag{3.45}$$

where as before

$$\mathbf{S}_p^{\tilde{y}} = \{\emptyset\}, \quad \mathbf{S}_{t-1}^{\tilde{y}} = (S_{t-1}^{\tilde{y}} = s_{t-1}^{\tilde{y}}, ..., S_{t-p}^{\tilde{y}} = s_{t-p}^{\tilde{y}}), \quad \text{for} \quad t \geq p+1$$

and

$$P[S_{p+1}^{\tilde{y}} = s_{p+1}^{\tilde{y}} \mid \mathbf{S}_p^{\tilde{y}}, Y, X] = P[S_{p+1}^{\tilde{y}} = s_{p+1}^{\tilde{y}} \mid Y, X],$$

such that $P_{\beta_1}[S_t^{\tilde{y}} = s_t^{\tilde{y}} \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X]$ and $P_0[S_t^{\tilde{y}} = s_t^{\tilde{y}} \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y]$ are in turn the p.m.fs of the signs for the transformed unrestricted and constrained regression models. Under the alternative hypothesis we have for $t = p, ..., T$

$$\log\left(P_{\beta_1}[S_t^{\tilde{y}} = s_t^{\tilde{y}} \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X]\right) = S_t^{\tilde{y}} \log P_{\beta_1}[\tilde{y}_t \geq 0 \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X] + (1 - S_t^{\tilde{y}}) \log P_{\beta_1}[\tilde{y}_t < 0 \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X], \tag{3.46}$$

which can be expressed as

$$\log\left(P_{\beta_1}[S_t^{\tilde{y}} = s_t^{\tilde{y}} \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X]\right) = S_t^{\tilde{y}} \log\left\{\frac{P_{\beta_1}[\tilde{y}_t \geq 0 \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X]}{P_{\beta_1}[\tilde{y}_t < 0 \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X]}\right\} + \log P_{\beta_1}[\tilde{y}_t < 0 \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X]. \tag{3.47}$$

Since under the null hypothesis the signs $S_t^{\tilde{y}}$, for $p \leq t \leq T$, are independent for the transformed regression (see Theorem 2), then

$$P[S_t^{\tilde{y}} = 1 \mid Y] = P[S_t^{\tilde{y}} = 0 \mid Y] = \frac{1}{2}, \quad t = p, ..., T.$$

Therefore, the sign-based Granger causality measure (3.45) can be expressed in the following manner

$$\tilde{C}(X \to Y) = \frac{1}{T-p+1} \sum_{t=p}^{T} \left\{ S_t^{\tilde{y}} \log \left\{ \frac{P_{\beta_1}[\tilde{y}_t \geq 0 \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X]}{P_{\beta_1}[\tilde{y}_t < 0 \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X]} \right\} + \log P_{\beta_1}[\tilde{y}_t < 0 \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X] - \log \left\{ \frac{1}{2} \right\} \right\} \tag{3.48}$$

Proposition 6 *Under assumptions (3.38) and (3.39)), let $H_0^{x \to y}$ and $H_1^{x \to y}$ be defined by (3.43) - (3.44),*

$$SG_T(A, \beta_1) = \sum_{t=p}^{T} S_t^{\tilde{y}} w_t(\beta_1), \tag{3.49}$$

where for $t = p, ..., T$

$$w_t(\beta_1) = \log \left\{ \frac{P_{\beta_1}[\tilde{y}_t \geq 0 \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X]}{P_{\beta_1}[\tilde{y}_t < 0 \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X]} \right\},$$

and suppose for any $0 \leq \alpha \leq 1$, $c_1(\alpha, \beta_1)$ is the critical value of the corresponding test with nominal level α, where $c_1(\alpha, \beta_1)$ is the smallest point such that

$$P\left[SG_T(A, \beta_1) > c_1(\alpha, \beta_1) \mid H_0^{x \to y}\right] \leq \alpha. \tag{3.50}$$

Then the test that rejects $H_0^{x \to y}$ when

$$SG_T(A, \beta_1) > c_1(\alpha, \beta_1)$$

is most powerful for testing $H_0^{x \to y}$ against $H_1^{x \to y}$ among level-α tests based on the signs $(S_p^{\tilde{y}}, ..., S_T^{\tilde{y}})$.

Under the null hypothesis of Granger non-causality, $S_p^{\tilde{y}}, ..., S_T^{\tilde{y}}$ are i.i.d according to a Bernoulli $Bi(1, 0.5)$. Henceforth, the distribution of the test statistic depends on the weights $w_t(\beta_1)$ and does not involve any nuisance parameter. Thus, it is distribution-free and robust against heteroskedasticity of unknown form, or as noted by Dufour and Taamouti (2010a) it is a *pivotal function* of nonparametric nature. On the other hand, under the alternative hypothesis, the power function depends on the form of the distribution of ε_t.

As it had been discussed in Section 3.3.1, the calculation of $w_t(\beta_1)$ depends on the joint distribution of process of signs which is computationally infeasible. Therefore, an assumption on the dependence structure of the process of signs is needed.

Assumption A5: *Let the signs $\{S_t^{\tilde{y}}\}_{t=p}^{\infty}$ follow a Markov process of order one. Then*

$$\begin{cases} P[\tilde{y}_t \geq 0 \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X] = P[\tilde{y}_t \geq 0 \mid \tilde{y}_{t-1} \geq 0, Y, X]^{S_{t-1}^{\tilde{y}}} P[\tilde{y}_t \geq 0 \mid \tilde{y}_{t-1} < 0, Y, X]^{1-S_{t-1}^{\tilde{y}}} \\ P[\tilde{y}_t < 0 \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X] = P[\tilde{y}_t < 0 \mid \tilde{y}_{t-1} \geq 0, Y, X]^{S_{t-1}^{\tilde{y}}} P[\tilde{y}_t < 0 \mid \tilde{y}_{t-1} < 0, Y, X]^{1-S_{t-1}^{\tilde{y}}} \end{cases}$$

Assumption A5 simplifies the calculations of the the probabilities $P[\tilde{y}_t \geq 0 \mid \mathbf{S}_{t-1}^{\tilde{y}}, X]$ and $P[\tilde{y}_t < 0 \mid \mathbf{S}_{t-1}^{\tilde{y}}, Y, X]$, and in turn, the calculation of the test statistic $SG_T(A, \beta_1)$.

Corollary 6 *Under the assumptions (3.38) and (3.39), and given transformation (3.40), the sign-based test of Granger non-causality from X to Y (3.49), rejects the null-hypothesis $\tilde{H}_0^{x \to y}$ when*

$$\widetilde{SG}_T(A, \beta_1) = \sum_{t=p}^{T} S_t^{\tilde{y}} S_{t-1}^{\tilde{y}} \tilde{\alpha}_t(\beta_1) + \sum_{t=p}^{T} S_t^{\tilde{y}} \tilde{w}_t(\beta_1) > c_1(\beta_1), \tag{3.51}$$

where

$$\tilde{w}_p(\beta_1) = \log\left\{\frac{1 - P[\varepsilon_p < -\beta_1' \underline{x}_0 \mid Y, X]}{P[\varepsilon_p < -\beta_1' \underline{x}_0 \mid Y, X]}\right\}, \quad \tilde{\alpha}_p(\beta_1) = 0$$

and for $t = p+1, ..., T$ we have

$$\tilde{w}_t(\beta_1) = \ln\left\{\frac{1 - \frac{P[\varepsilon_t < -\beta_1' \underline{x}_{t-1},\ \varepsilon_{t-1} < -\beta_1' \underline{x}_{t-2} | Y, X]}{P\left[\varepsilon_{t-1} < -\beta_1' \underline{x}_{t-2} | Y, X\right]}}{\frac{P\left[\varepsilon_t < -\beta_1' \underline{x}_{t-1},\ \varepsilon_{t-1} < -\beta_1' \underline{x}_{t-2} | Y, X\right]}{P\left[\varepsilon_{t-1} < -\beta_1' \underline{x}_{t-2} | Y, X\right]}}\right\}$$

$$\tilde{\alpha}_t(\beta_1) = \ln\left\{\frac{1 - \left(\frac{P[\varepsilon_t < -\beta_1' \underline{x}_{t-1} | Y, X]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' \underline{x}_{t-2} | Y, X\right]} - \frac{P[\ \varepsilon_{t-1} < -\beta_1' \underline{x}_{t-2},\ \varepsilon_t < -\beta_1' \underline{x}_{t-1} | Y, X]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' \underline{x}_{t-2} | Y, X\right]}\right)}{\frac{P\left[\varepsilon_t < -\beta_1' \underline{x}_{t-1} | Y, X\right]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' \underline{x}_{t-2} | Y, X\right]} - \frac{P\left[\ \varepsilon_{t-1} < -\beta_1' \underline{x}_{t-2},\ \varepsilon_t < -\beta_1' \underline{x}_{t-1} | Y, X\right]}{1 - P\left[\varepsilon_{t-1} < -\beta_1' \underline{x}_{t-2} | Y, X\right]}}\right\} - \ln\left\{\frac{1 - \frac{P[\varepsilon_t < -\beta_1' x_{t-1},\ \varepsilon_{t-1} < -\beta_1' \underline{x}_{t-2} | Y, X]}{P\left[\varepsilon_{t-1} < -\beta_1' \underline{x}_{t-2} | Y, X\right]}}{\frac{P\left[\varepsilon_t < -\beta_1' \underline{x}_{t-1},\ \varepsilon_{t-1} < -\beta_1' \underline{x}_{t-2} | Y, X\right]}{P\left[\varepsilon_{t-1} < -\beta_1' \underline{x}_{t-2} | Y, X\right]}}\right\}$$

The test statistic $\widetilde{SG}_T(A, \beta_1)$ can be simulated under the null hypothesis and the relevant critical values can be estimated with sufficient number of replications [see Dufour and Taamouti (2010a)]. The confidence set with level $(1 - \alpha)$, are set of values β_0, for which we cannot reject the null hypothesis of Granger non-causality. In other words, the confidence set for β_0 can be expressed as follows

$$J(\alpha, A) = \left\{\beta_0 : \widetilde{SG}_T(A, \beta_1) \leq c_1(\alpha, \beta_1)\right\} \tag{3.52}$$

where by construction

$$P\left[\beta \in J(\alpha, A)\right] \geq 1 - \alpha \tag{3.53}$$

3.4.2 Inference with *unknown* nuisance parameters

If the nuisance vector A is unknown, one natural approach for dealing with the nuisance parameter problem may appear to be the estimation of the unknown coefficients $\hat{A}$. Then given the estimated value, we may construct Granger non-casuality tests using transformation (3.40) and considering the test statistic $\widetilde{SG}_T(\hat{A}, \beta_1)$ - this apporach has been considered in the simulations section. However, provably valid procedures for testing Granger non-causality in the presence of unknown nuisance parameters can be obtained using the simultaneous inference approach implemented by Dufour (1990) and Campbell and Dufour (1997) [see Cavanagh et al. (1995) and Campbell and Yogo (2006) among others for use in different settings]. This is accomplished as follows: first we construct an exact $(1-\alpha_1)$ confidence set, say $CS_A(\alpha_1)$, for the vector of nuisance parameters, A, such that $0 \leq \alpha_1 \leq \alpha \leq 1$ and

$$P\left[A \in CS_A(\alpha_1)\right] \geq 1 - \alpha_1. \tag{3.54}$$

Second, as shown before, given the *true* value of A, it is possible to obtain an exact confidence set for β. Therefore, once a confidence set for A is available, sign-based Granger non-causality tests corresponding to each value A in the confidence set $CS_A(\alpha_1)$ are constructed at level α_2 (i.e. $\widetilde{SG}(A, \beta_1; \alpha_2)$), where $0 \leq \alpha_2 \leq 1$. Finally, the conditional sign-based tests $\widetilde{SG}(A, \beta_1; \alpha_2)$ and the confidence set for A are combined to obtain a simultaneous confidence set for A and β. Then an intersection-union method is used to obtain confidence sets and valid tests for β, irrespective of the true value of A.

The confidence set for A can be constructed in a variety of different ways; however, preference goes to more powerful testing procedures, as they result in a tighter confidence set [see Campbell and

Yogo (2006)]. The issue of constructing the confidence set for the nuisance vector A is addressed later on in this Section. In what follows, we first consider the case of constructing a simultaneous confidence set for A and β, before turning our attention to making inference only on β. In what follows, we follow Dufour (1990) by first considering the problem of constructing a confidence set for (A, β), which in turn allows us to test joint hypotheses of the form $\bar{H}_0 : A = A_0, \beta = \beta_0$.

Let us co+struct a confidence set for (A, β), say $K(\alpha_1, \alpha_2)$, with $0 \leq \alpha_1, \alpha_2 \leq 1$, such that

$$\begin{aligned} K(\alpha_1, \alpha_2) &= \{(A, \beta) : A \in CS_A(\alpha_1) \cap \beta \in J(\alpha_2, A)\} && (3.55) \\ &= \left\{(A, \beta) : A \in CS_A(\alpha_1) \cap \widetilde{SG}_T(A, \beta_1) \leq c_1(\alpha_2, \beta_1)\right\}. && (3.56) \end{aligned}$$

which naturally implies that

$$P\left[(A, \beta) \in K(\alpha_1, \alpha_2)\right] = P\left[A \in CS_A(\alpha_1) \cap \widetilde{SG}_T(A, \beta_1) \leq c_1(\alpha_2, \beta_1)\right]. \tag{3.57}$$

Finally, by employing De Morgan's law and Bonferroni inequality, we obtain

$$\begin{aligned} P\left[(A, \beta) \in K(\alpha_1, \alpha_2)\right] &= 1 - P\left[A \notin CS_A(\alpha_1) \cup \beta \notin J(\alpha_1, A)\right] \\ &= 1 - P\left[A \notin CS_A(\alpha_1) \cup \widetilde{SG}_T(A, \beta_1) > c_1(\alpha_2, \beta_1)\right] \\ &\geq 1 - P[A \notin CS_A(\alpha_1)] - P[\widetilde{SG}_T(A, \beta_1) > c_1(\alpha_2, \beta_1)] \\ &\geq 1 - \alpha_1 - \alpha_2, \end{aligned} \tag{3.58}$$

where $K(\alpha_1, \alpha_2)$ is a confidence set for (A, β) with level $(1 - \alpha) \equiv 1 - \alpha_1 - \alpha_2$. The values of α_1 and α_2 can be chosen such that the desired level α is achieved. With the setup above, it is clear that the joint null hypothesis $\bar{H}_0 : A = A_0, \beta = \beta_0$ gets rejected when $(A_0, \beta_0) \notin K(\alpha_1, \alpha_2)$. In turn, given (3.58), this would imply

$$P\left[(A, \beta) \notin K(\alpha_1, \alpha_2)\right] \leq \alpha_1 + \alpha_2 \equiv \alpha,$$

which is the desired level α.

The issue with the earlier result is that it considers testing the joint null hypothesis $\bar{H}_0 : A =$

$A_0, \beta = \beta_0$, as opposed to only testing the null hypothesis of Granger non-causality (i.e. $\tilde{H}_0^{x \to y}$: $\beta = 0$). In other words, we obtained simultaneous confidence set $K(\alpha_1, \alpha_2)$, where the test is only rejected when $(A_0, \beta_0) \notin K(\alpha_1, \alpha_2)$; although, conservative confidence sets for the vector β_0 can then be obtained using the projection technique.

To make inference only on the parameter vector β, we adapt Proposition 2 of Dufour (1990) to the context of our study. To do this, we first introduce some statistical terminology: Let $\underset{A \in CS_A(\alpha_1)}{\widetilde{SG}}(A, \beta_1) \in \mathcal{S}$, where $\mathcal{S} \subset \mathbb{R}$. Denote $(\mathcal{R}, \mathcal{A})$ as paritions of $\mathcal{S}$, such that if $\underset{A \in CS_A(\alpha_1)}{\widetilde{SG}}(A, \beta_1) \in \mathcal{R}$, the null hypothesis $H_0^{x \to y}$ is rejected, and if $\underset{A \in CS_A(\alpha_1)}{\widetilde{SG}}(A, \beta_1) \in \mathcal{A}$, the null hypothesis is accepted. A test for the null hypothesis $H_0^{x \to y}$: $\beta = 0$ with nominal level α is conservative if $P\left[\underset{A \in CS_A(\alpha_1)}{\widetilde{SG}}(A, \beta_1) \in \mathcal{R} \mid \beta = 0\right] \leq \alpha$, and it is liberal if $P\left[\underset{A \in CS_A(\alpha_1)}{\widetilde{SG}}(A, \beta_1) \in \mathcal{R} \mid \beta = 0\right] \geq \alpha$. For a conserative test if the critical region is expanded to its nominal level α, the conclusion of the test remains the same. Similarly, for a liberal test if the critical region is contracted to its nominal level, the conclusion is unchanged. Now if for a null hypothesis $H_0^{x \to y}$, we can construct both liberal and conservative tests with partitions $(\mathcal{R}_1, \mathcal{A}_1)$ and $(\mathcal{R}_2, \mathcal{A}_2)$ respectively, such that $\mathcal{R}_1 \subseteq \mathcal{R}_2$, then the test is rejected if $\underset{A \in CS_A(\alpha_1)}{\widetilde{SG}}(A, \beta_1) \in \mathcal{R}_1$ and accepted if $\underset{A \in CS_A(\alpha_1)}{\widetilde{SG}}(A, \beta_1) \in \mathcal{A}_2$; otherwise, if $\underset{A \in CS_A(\alpha_1)}{\widetilde{SG}}(A, \beta_1) \in \mathcal{R}_2 - \mathcal{R}_1$, both tests are inconlusive. Hence, unlike the "traditional bound-type procedures" which rely only on one test-statistic, tests surrounding parameter vector β are based on two test statistics with nested critical regions, such that the smaller and the larger regions in turn yield conservative and liberal tests, with the difference between the two regions being considered as an inconclusive region. Corresponding to the conservative and libral tests with nominal level α for the null hypothesis $\tilde{H}_0^{x \to y}$: $\beta = 0$, we may also construct conservative and liberal confidence sets for β, such that $P[\beta \in U] \geq 1 - \alpha$ and $P[\beta \in L] \leq 1 - \alpha$, where U and L are in turn conservative and liberal confidence sets.

According to the earlier definitions, inference on the null hypothesis $\tilde{H}_0^{x \to y}$: $\beta = 0$ with nominal level α, with $0 \leq \alpha \leq 1$, requires the construction of conservartive and liberal tests with a nested critical region, such that paritions R_1 and R_2 in $\mathcal{S}$ satisfy $R_1 \subseteq R_2$. For this to hold, the conservative test must have level α_2 with $0 \leq \alpha_2 \leq \alpha$, while the liberal test must have level α_3, with $\alpha_3 \geq \alpha \geq \alpha_2 \geq 0$. Thus, consider model (3.40) and let $CS_A(\alpha_1)$ be the confidence set for A.

Following Dufour (1990), α_1, α_2 and α_3 are chosen such that

$$\alpha_1 = \alpha - \alpha_2, \quad \alpha_3 = \alpha + \alpha_1, \quad 0 \leq \alpha_1 < \alpha \leq 1 - \alpha_1. \tag{3.59}$$

In turn, conservative and liberal sets for β are defined as

$$U(\alpha_1, \alpha_2) = \{\beta : (A, \beta) \in K(\alpha_1, \alpha_2), \text{ for some } A \in CS_A(\alpha_1)\}, \tag{3.60}$$

$$L(\alpha_1, \alpha_3) = \{\beta : (A, \beta) \in K(\alpha_1, \alpha_2), \forall A \in CS_A(\alpha_1)\}. \tag{3.61}$$

If (3.59) holds, $U(\alpha_1, \alpha_2)$ and $L(\alpha_1, \alpha_3)$ are in turn conservative and liberal sets for β with the same level $1 - \alpha$, such that $P[\beta \in U(\alpha_1, \alpha_2)] \geq 1 - \alpha$ and $P[\beta \in L(\alpha_1, \alpha_3)] \leq 1 - \alpha$ [see proof in Appendix A.1. of Dufour (1990)]. Therefore, with these sets, conservative and liberal tests can be constructed. Let

$$Q_L(\beta) = \inf \left\{ \widetilde{SG}_T(A, \beta_1) : A \in CS_A(\alpha_1) \right\}, \tag{3.62}$$

$$Q_U(\beta) = \sup \left\{ \widetilde{SG}_T(A, \beta_1) : A \in CS_A(\alpha_1) \right\}. \tag{3.63}$$

Then $\beta \notin U(\alpha_1, \alpha_2)$ is equivalent to $Q_L(\beta) \geq S(\alpha_2)$, while $\beta \in L(\alpha_1, \alpha_3)$ is equivalent to $Q_U(\beta) \leq S(\alpha_3)$. Thus, the bounds test for testing $\tilde{H}_0^{x \to y} : \beta = 0$

$$\begin{cases} \text{Rejects } \tilde{H}_0^{x \to y} \text{ when } Q_L(\beta) > c_1(\alpha_2, \beta_1), \\ \text{Accepts } \tilde{H}_0^{x \to y} \text{ when } Q_U(\beta) \leq c_1(\alpha_3, \beta_1), \\ \text{Inconclusive otherwise.} \end{cases} \tag{3.64}$$

The issue of choosing the value of α_1, and hence α_2 and α_3 is discussed in the next Section.

To find an exact confidence set for the vector of unknown nuisance parameters, which must at least be true under the null hypothesis of Granger non-causality [see. Campbell and Dufour (1997)], we take advantage of the exact estimation procedure proposed by Andrews (1993), based on median-bias correction for AR(1) models with an intercept. The extension to AR(p) processes can be achieved with slight modifications [see. Andrews and Chen (1994), Rudebusch (1992), and

Stock (1991)]. However, the adjusted methodology will no longer yield exact results, but rather *"approximate"* confidence intervals.

For the simplicity of exposition let us consider a process with lag order $p = 1$. Then we can express the constrained regression equation (3.37) under the null hypothesis of Granger non-causality as follows

$$y_t = m_1 + a_1 y_{t-1} + \varepsilon_t, \quad t = 1, ..., T \tag{3.65}$$

with $a_1 \in (-1, 1]$, and where it is now assumed that ε_t are independent and identically distributed error terms, such that $\varepsilon_t \sim N(0, \sigma^2)$. If $\mid a_1 \mid < 1$, then an assumption regarding the initial condition is made -e.g. $y_0 \sim N(m_1, \sigma^2/(1 - a_1{}^2))$; otherwise, when $a_1 = 1$ and the process is non-stationary, y_0 has an arbitrary initial condition. Andrews (1993), suggests that in the presence of an intercept (such as m_1), the least squares estimator of parameter a_1 in the AR(1) model (3.65) has downward bias, particularly when a_1 is close to unity. Therefore, he proposes a bias-correction strategy based on the median-bias (i.e. the difference between the median of the estimator and its true value) of the LS estimator of a_1, where the same strategy can also be utilized to obtain exact confidence intervals. The estimator $\hat{a}_1$ is median-unbiased, when the true parameter a_1 is the median of $\hat{a}_1$. In other words,

$$\mathbb{E}_{a_1} \mid \hat{a}_1 - \alpha_1 \mid \leq \mathbb{E}_{a_1} \mid \hat{a}_1 - a_1' \mid, \quad \forall a_1, a_1' \in \Theta, \tag{3.66}$$

where Θ is the parameter space for a_1. Property (3.66) implies that on average the distance between $\hat{a}_1$ and the true parameter value a_1, is less than the distance between $\hat{a}_1$ and any other parameter. Let $q_p(a_1)$ denote the p^{th} quantile function of $\hat{a}_1$, where the distribution of $\hat{a}_1$ depends only on a_1 [see appendix A of Andrews (1993) for the proof of the invariance property of $\hat{a}_1$]. By definition, for $p \in (0, 1)$, $P\left[\hat{a}_1 \leq q_p(a_1)\right] = p$. The $(1 - p)$ confidence interval for a_1 is given by the set

$$\left\{a_1 \in [-1, 1] : q_{p_1}(a_1) \leq \hat{a}_1 \leq q_{p_2}(a_1)\right\}, \tag{3.67}$$

where $p_1 \geq 0$, $p_2 \geq 0$ and since

$$\begin{aligned} P\left[q_{p_1}(a_1) \leq \hat{a}_1 \leq q_{p_2}(a_1)\right] &= P\left[\hat{a}_1 \leq q_{p_2}(a_1)\right] - P\left[\hat{a}_1 \leq q_{p_1}(a_1)\right] \\ &= (p_2 - p_1) \\ &= 1 - p, \end{aligned}$$

it is evident that $p = 1 + p_1 - p_2$. Quantile functions $q_{p_1}(a_1)$ and $q_{p_2}(a_1)$ correspond to interval $\{a_1 : \widehat{L} \leq a_1 \leq \widehat{U}\}$ as shown in Andrews (1993), they are both increasing in a_1. Andrews (1993) defines $\widehat{L}$ and $\widehat{U}$ as follows:

$$\widehat{L} = \begin{cases} > 1 & \text{if} \quad \hat{a}_1 > q_{p_2}(1) \\ q_{p_2}^{-1}(\hat{a}_1) & \text{if} \quad q(-1) < \hat{a_1} \leq q(1) \\ -1 & \text{if} \quad \hat{a}_1 \leq q_{p_2}(-1) \end{cases}, \quad \widehat{U} = \begin{cases} > 1 & \text{if} \quad \hat{a}_1 > q_{p_1}(1) \\ q_{p_1}^{-1}(\hat{a}_1) & \text{if} \quad q(-1) < \hat{a_1} \leq q(1) \\ -1 & \text{if} \quad \hat{a}_1 \leq q_{p_1}(-1) \end{cases}, \tag{3.68}$$

where by definition for $j = 1, 2$, $q_{p_j}(-1) = \lim_{\alpha_1 \to -1} q_{p_j}(a_1)$ and $q_{p_j}^{-1} : (q_{p_j}(-1), q_{p_j}(1)] \to (-1, 1]$ is the inverse function of $q_{p_j}(.)$, such that $q_{p_j}^{-1}\left(q_{p_j}(a_1)\right) = a_1$.

Evidently, the shortcoming of this procedure is that for the interval to be exact, the distribution of the residuals must be specified as normal. However, Andrews (1993) provides clear numerical evidence that the aforementioned procedures based on normal distribution are robust against non-normal distributions - results which are reaffirmed by our simulations exercise in Section 3.5.

3.5 Monte Carlo simulations

In this Section, we assess the finite sample bias of the estimator of the proposed sign-based Granger causality measures and recommend a VAR Sieve bootstrap procedure to estimate the bias-corrected causality measures. Thereafter, we examine the performance of the sign-based Granger non-causality tests in finite samples. To assess the empirical size and power of the tests, we consider a variety of different DGPs which correspond to symmetric and asymmetric distributions often encountered in practice.

3.5.1 Bootstrap bias-corrected estimation of sign-based causality measures

In theory it may be possible to derive analytical expressions for the sign-based Granger causality measures. For instance, in the case of Granger non-causality from X to Y, it is immediately evident from the law of large numbers that $\hat{C}_T(X \to Y) \to 0$ as $T \to \infty$. However, in more complex cases, theoretical derivation of the causality measure values may not be practical. As noted by Dufour and Taamouti (2010b), the root of the bias in autoregressive coefficients is finite sample sizes. Thus, where theoretical derivations are infeasible, the causality measure values can be simulated. The idea consists of simulating a large sample from an unrestricted model with known parameters, and in turn using the simulated sample to estimate the parameters of the constrained (i.e. in the case of Granger non-causality) model. The *large simulation* algorithm of Dufour and Taamouti (2010b) is tailored to fit the sign-based Granger causality measures proposed in our study as follows:

1. Simulate a large sample of T observations using the unrestricted model with known parameters and initial values, under the assumption that the distribution of the residuals ε_t is specified. As it will be shown in Section 3.5.1.1, and as noted by Dufour and Taamouti (2010b), the choice of the distribution of the residuals does not have a significant impact on the value of the causality measures. In our study, we consider a sample size of $T = 1,000,000$; however, Dufour and Taamouti (2010b) have shown that sample sizes as small as $600,000$ yield results consistent with the theoretical values of the causality measures.

2. Use the simulated large sample to estimate the parameters of the constrained model.

3. Obtain the signs S_t^y, such that $S_t^y = \mathbb{1}_{\mathbb{R}^+ \cup \{0\}}\{y_t\}$, for $t = 0, ..., T$, where $\mathbb{1}$ is an indicator function, and calculate the CDFs and the joint CDFs of the residuals $\{\varepsilon_t\}_{t=1}^T$ corresponding to the unrestricted and constrained regressions as shown in Section 3.3.2.

4. Calculate the sign-based Granger causality measure derived in corollary 5.

To reduce bias, we follow Dufour and Taamouti (2010b) and Taamouti et al. (2014) and use bootstrap to compute the small sample bias in the estimator of the sign-based Granger causality measures. The bias-term is then subtracted to obtain the bootstrap bias-corrected estimates. Taamouti et al. (2014) approximate the bootstrap bias term $Bias = \mathbb{E}\left[\hat{C}(X \to Y) - C(X \to Y)\right]$ by

$$Bias_{boot} = \mathbb{E}_{boot}\left[\hat{C}_{boot}(X \to Y) - \hat{C}(X \to Y)\right], \tag{3.69}$$

where $\mathbb{E}_{boot}$ is the expectation with respect to the distribution of the bootstrap sample $\hat{C}_{boot}(X \to Y)$, and $\hat{C}(X \to Y)$ is the estimate of the sign-based causality measure using the original sample. In practice, the expectation operator of the bias estimator (3.69) can be replaced with its empirical counterpart, - i.e. the sample mean. This implies that the bias term is estimated as follows

$$\widehat{Bias}_{boot} = \frac{1}{B}\sum_{k=1}^{B} \hat{C}_{boot}^{(k)}(X \to Y) - \hat{C}(X \to Y). \tag{3.70}$$

A bootstrap procedure based on simple sampling with replacement does not preserve the conditional dependence structure of the data. Therefore, we suggest the vector autoregressive sieve bootstrap procedure proposed by Meyer and Kreiss (2015), which is simply an extension of the AR sieve bootstrap for multivariate time-series. In other words, this bootstrap procedure fits a VAR process to the data, as opposed to a an AR process. To obtain the asymptotic validity of the VAR sieve bootstrap, the conditions outlined in Theorem 4.1 of Meyer and Kreiss (2015) must hold - these conditions are discussed in detail in the Appendix. The process of estimating the bias using the vector autoregressive sieve bootstrap is then as follows

1. Select an order p of a VAR model for the bivariate vector $\{Z_t : t = 0, 1, ..., T\}$ such that $p << T$. An appropriate lag order p can be selected using a criterion, such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC).

2. Fit the vector autoregressive model to the simulated observations using the multivariate least squares (LS) or the Yule-Walker estimator. We follow Meyer and Kreiss (2015) and Bühlmann et al. (1997) by estimating the parameters using the latter approach, where the parameter estimates for the VAR(p) model are denoted by $\hat{\Phi}_1(p), ..., \hat{\Phi}_p(p)$ and are the

solution to the following linear system

$$\left(\hat{\Phi}_1(p), ..., \hat{\Phi}_p(p)\right) \hat{G}(p) = \left(\hat{\Gamma}(1), ..., \hat{\Gamma}(p)\right)$$

where for each lag $-p \leq \tau \leq p$, $\hat{\Gamma}(\tau)$ is the sample autocovariance matrix of $Z_0, ..., Z_T$ and $\hat{G}(p) \in \mathbb{R}^{2p \times 2p}$ is defined by $\hat{G}(p) = \left(\hat{\Gamma}(s-r)\right)_{r,s=1,...,p}$.

3. Let $\varepsilon'_t = Z_t - \hat{c} - \sum\limits_{j=1}^{p} \hat{\Phi}_j(p) Z_{t-j}$, $t = p, ..., T$ be the underlying residuals of the fitted vector autoregressive model and $\hat{F}_T$ be the empirical distribution function of the centered residuals $\hat{\varepsilon}_t = \varepsilon'_t - \bar{\varepsilon}$, where $\bar{\varepsilon} = (T-p+1)^{-1} \sum_{t=p}^{T} \varepsilon'_t$.

4. Generate $T + p$ independent residuals $\varepsilon^*_1, \varepsilon^*_2, ..., \varepsilon^*_{T+p}$ from the distribution $\hat{F}_T$ by random sampling with replacement. Using these generated residuals ε^*_t along with the Yule-Walker coefficient estimates $\hat{\Phi}$, we generate a bootstrap sample $(Z^*_p, ..., Z^*_T)$ according to the equation

$$Z^*_t = \sum_{j=1}^{p} \hat{\Phi}_j(p) Z^*_{t-j} + \varepsilon^*_t,$$

The first p data points are later discarded.

5. Estimate the least-squares estimates of the unconstrained and constrained marginal processes y^*_t and x^*_t based on the pseudo time-series $Z^*_1, ..., Z^*_T$.

6. Calculate the causality measure $\hat{C}^{(k)}_{boot}(X \rightarrow Y)$ based on the pseudo time-series $Z^*_1, ..., Z^*_T$.

7. Repeat steps $2 - 6$, B times.

8. Calculate the bias term $\widehat{bias}_{boot}$, using relationship (3.70).

9. Calculate the bias-corrected estimate of the sign-based Granger causality measure as follows:

$$\hat{C}_{BC}(X \rightarrow Y) = \hat{C}(X \rightarrow Y) - \widehat{Bias}_{boot} \tag{3.71}$$

Since in practice the bias-corrected causality measure $\hat{C}_{BC}(X \rightarrow Y)$ can be negative for some bootstrap samples, we follow Taamouti et al. (2014), by imposing the following non-negativity truncation

$$\hat{C}_{BC}(X \to Y) = \max\left\{\hat{C}_{BC}(X \to Y), 0\right\} \tag{3.72}$$

The bias-corrected sign-based causality measure from Y to X can be estimated in the same manner. It is crucial to point out that the bias-corrected estimator of sign-based causality measures is a raw measure of causality, in the sense that while it is useful for comparing and assessing the strength of the causal relationship, hypotheses tests cannot be made on the value of the measure itself to infer on its significance. Therefore, in what follows, we show that by performing slight modifications to the sign-based measures, Granger non-causality tests can be constructed to assess the presence of a causal relationship or lack thereof. The Bias-corrected estimators of sign-based measures in conjunction with the sign-based Granger non-causality tests equip us with the means of both assessing the strength of the causal relationship and testing for its presence.

3.5.1.1 Simulation study

The aim of this Section is to run a Monte Carlo experiment to investigate the possible bias in the estimator of the sign-based Granger causality measures proposed in Section 3.3.2. The data generating processes considered represent linear regression models under different distributional assumptions and forms of heteroskedasticity. Table 3.1 presents the DGPs that have been utilized in our simulation study, with the last column showing the direction of causality. First let us consider DGP1: it is evident that by construction X and Y are independent - i.e. there is no causality from X to Y, nor from Y to X. Therefore, it can be deduced that the sign-based causality measures possess known true values of zero, and as such $C(X \to Y) = C(Y \to X) = 0$. The same goes for DGP2 in which X and Y are independent conditional on their past values. These two processes are therefore good benchmarks by which the bias in the estimators of the sign-based causality measures can be evaluated. On the other hand, the case for Granger non-causality from X to Y is not present in DGPs 3 and 4, which in turn are of unidirectional and bidirectional causal natures. In DGP3, for instance, X Granger causes Y, while Y is independent of X conditional on the past value of X. Therefore, in this case the true measure of causality from X to Y is unknown and non-zero. Finally referring to DGP4, it is evident that the causality relationship between Y and X is bidirectional, where in addition X and Y exhibit simultaneous causality. Despite the fact that the true values of the sign-based Granger causality measures from X to Y is unknown for DGPs 3 and 4, we still have the means of evaluating the finite sample bias of the estimators by employing the *large simulation* algorithm proposed by Dufour and Taamouti (2010b), introduced in Section 3.5.1. We have chosen $T = 1,000,000$ for the aforementioned algorithm, which is significantly greater than the minimum sample size (i.e. 600,000) proposed by Dufour and Taamouti (2010b); this is to ensure that the sample size related coefficient bias is minimized. In addition to DGPs 1 to 4, which all possess bivariate normally distributed disturbances, we have further considered the marginal processes of DGP3 under different distributional assumptions and forms of heteroskedasticity [see table 3.2]. As it has been noted by Dufour and Taamouti (2010b), different distributional assumptions should not lead to different values of causality measures, and these additional assumptions are imposed merely for assessing the robustness of the exact Granger non-casuality test. Nevertheless, for the sake of rigor, the bias in the estimators of the sign-based Granger causality meausres for DGP3(I)-(VIII) have been considered and evaluated as well.

We consider sample sizes of $T = 50, 150, 250$ and 500 to simulate the bias-corrected estimators of the sign-based Granger causality measures, and $T = 50$ and 150 to examine the finite sample properties of the exact sign-based inference procedure in terms of size and power. The smaller sample sizes for the latter exercise is due to limitations in computational power, since the Imhof (1961) algorithm used for finding the exact confidence set for the vector of nuisance parameters performs very slowly in samples greater than 100 observations. Tables 3.3 and 3.4 present the results for the bootstrap bias-corrected estimators of the sign-based Granger causality measures for different sample sizes. The bias term $\widehat{Bias}_{boot}$ used for calculating the bias-corrected measures has been estimated by considering $B = 500$ sample estimates of the causality measure based on the VAR sieve bootstrapped data - these are presented in rows labeled 'Bias corrected', with their corresponding standard errors shown in brackets. On the other hand, the rows labeled 'True' present the values of the causal measures that are known, either due to clearly evident independence (conditional independence) of the variables X and Y, or because they have been

Table 3.1: Data-generating processes (DGPs) considered in the Monte Carlo study to assess the finite sample bias of the estimator of the sign-based Granger causality measures

DGPs	Variables y_t	Variables x_t	Direction of Causality
DGP1	ε_t^y {white noise}	ε_t^x {white noise}	$Y \nrightarrow X, X \nrightarrow Y$
DGP2	$y_t = 0.5y_{t-1} + \varepsilon_t^y$	$x_t = 0.5x_{t-1} + \varepsilon_t^x$	$Y \nrightarrow X, X \nrightarrow Y$
DGP3(I-VIII)	$y_t = 0.5y_{t-1} + 0.5x_{t-1} + \varepsilon_t^y$	$x_t = 0.9x_{t-1} + \varepsilon_t^x$	$Y \nrightarrow X, X \rightarrow Y$
DGP4	$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 0.2 \\ 0.3 \end{pmatrix} + \begin{bmatrix} 0.1 & 0.8 \\ 0.7 & 0.15 \end{bmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^y \\ \varepsilon_t^x \end{pmatrix}$, with $\begin{pmatrix} \varepsilon_t^y \\ \varepsilon_t^x \end{pmatrix} \sim \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix} \right]$		$Y \rightarrow X, X \rightarrow Y$ $Y \leftrightarrow X$

Note: This table contains the different data generating processes considered in our study for investigating the bias in the estimation of the sign-based of Granger causality measures, as well as the empirical size and power of the Granger non-causality test. The process (y_t, x_t) is simulated for $t = 1, ..., T$ with the assumption that $(\varepsilon_t^y, \varepsilon_t^x)$ are $i.i.d$ from $N(0, \Sigma_2)$. DGP3(I)-DGP(VIII) further consider different distributional assumptions and forms of heteroskedasticiy.

simulated by employing the large simulation algorithm in cases where causal relationships are in fact present.

First, let us consider DGPs 1 and 2: it is evident that for the cases of Granger non-causality, the bias-corrected estimators of the sign-based causality measures are close to zero. Furthermore, the estimator appears to be consistent, since increases in sample size leads to the convergence of the bias-corrected estimator of the causality measure to its true value of zero. Similar phenomenon can be observed for DGPs 3(I-VIII) and DGP4, where cases of Granger causality from X to Y and Y to X are present. For instance, there is inherent causality from X to Y in DGP3(I), whereas X is independent of Y conditional on the past value of X. Therefore, the bias-corrected estimator of the sign-based Granger causality measure from X to Y is non-zero, and it approaches its simulated true value of 0.0771 as the sample size increases; however, due to the conditional independence of Y and X in DGP3(I), the estimator converges to zero. Under different distributional assumptions,

Table 3.2: Residuals of DGP3 with different distributional assumptions and forms of heteroskedasticity

DGP3		
I	$\varepsilon_t^y, \varepsilon_t^x \sim N(0,1)$	
II	$\varepsilon_t^y, \varepsilon_t^x \sim Cauchy$	
III	$\varepsilon_t^y, \varepsilon_t^x \sim t(2)$	
IV	$\varepsilon_t^y, \varepsilon_t^x \sim s_t \mid \varepsilon_t^C \mid -(1-s_t) \mid \varepsilon_t^N \mid$	where $P(s_t=1)=P(s_t=0)=\frac{1}{2}$
V	$\varepsilon_t^y, \varepsilon_t^x \sim \begin{cases} N(0,1) & \text{for } t \neq 25 \\ \sqrt{1000}N(0,1) & \text{for } t = 25 \end{cases}$	
VI	$\varepsilon_t^y, \varepsilon_t^x \sim N(0,\sigma_\varepsilon^2(t))$ and $\sigma_\varepsilon(t) = \exp(0.5t)$	
VII	$\varepsilon_t^y, \varepsilon_t^x \sim \begin{cases} N(0,\sigma_\varepsilon^2(t)) & \text{for } t \neq 25 \\ 50N(0,\sigma_\varepsilon^2(t)) & \text{for } t = 25 \end{cases}$	$\sigma_\varepsilon^2(t) = 0.00037 + 0.0888\varepsilon_{t-1}^2 + 0.9024\sigma_\varepsilon^2(t-1)$
VIII	$\varepsilon_t^y, \varepsilon_t^x \sim N(0,\sigma_\varepsilon^2(t))$	$\sigma_\varepsilon^2(t) = 0.45\varepsilon_{t-1}^2 + 0.45\sigma_\varepsilon^2(t-1)$

Note: This table summarizes different symmetric and asymmetric distributions and the forms of heteroskedasticity for the marginal processes Y and X of DGP3: (I) Normal distribution; (II) Cauchy distribution; (III) Student t distribution with two degrees of freedom; (IV) Mixture of Normal and Cauchy distributions where ε_t^C follows a Cauchy distribution, ε_t^N follows a $N(0,1)$ distribution; (V) Break in variance; (VI) exponential variance; (VII) GARCH(1,1) plus jump in variance; (VIII) GARCH(1,1) variance.

our results for DGP3(I-VIII) mostly agree with Dufour and Taamouti (2010), in that the values of the causality measures are not sensitive to the form of the distribution. However, this is evidently not the case for DGP3(II), which is generated by a Cauchy distribution (with an undefined mean), and DGP3(VIII) which is based on a normal distribution with GARCH(1,1).

3.5.2 Empirical size and power of Granger causality tests

In this Section, we assess the finite sample performance of the sign-based Granger non-causality tests introduced in Section 3.4. The empirical size and the power of the tests are examined using the DGPs in tables 3.1 and 3.2. The marginal processes Y in DGPs 1 and 2 and X in DGPs 1-3 are used to examine the size of the test, since by construction these satisfy the null hypothesis of Granger non-causality. On the other hand, marginal processes Y in DGPs 3 and 4, and X in DGP 4 are used to investigate the power of the sign-based Granger non-causality tests. Our main interest lies in examining the performance of the said tests within the domains of DGP3(I-VIII), as these data are generated under different distributional assumptions and forms of heteroskedasticy; hence, they would reveal whether the proposed tests are robust against non-normal and heteroskedastic distributions.

As noted earlier, the confidence set for the vector of nusiance parameters is obtained by finding the exact median-biased corrected quantiles using the Imhof (1961) algorithm. In order to keep the computation of the simulations within a reasonable time frame, only sample sizes of $T = 50, 150$ are considered, as the Imhof (1961) algorithm performs very slowly for samples of greater than 100 observations. The distribution of the test statistic of the transformed regression is simulated under the null hypothesis of Granger non-causality, where the critical values are then estimated to any degree of precision with sufficient number of replications. In our study, we considered 1,000 replications for simulating the distribution and 500 iterations for simulating the size and power of the test.

The implementation of the bound-type procedure suggested in Section 3.4.2 entails fixing an arbitrary significance level, say α, and choosing the width of the confidence set $CS_A(\alpha_1)$ for the vector of nuisance parameters such that the power of the test is maximized. Once α_1 is chosen, α_2

Table 3.3: Bootstrap bias-corrected estimators of the sign-based Granger causality measures for $T = 50$ and $T = 150$

	Measure	DGP1	DGP2	DGP3(I)	DGP3(II)	DGP3(III)	DGP3(IV)	DGP3(V)	DGP3(VI)	DGP3(VII)	DGP3(VIII)	DGP4
Sample size: $T = 50$												
$X \to Y$		No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
True	$C(X \to Y)$	0.0000	0.0000	0.0771	0.0385	0.0560	0.0738	0.0772	0.0769	0.0786	0.1061	0.1565
Bias corrected	$\hat{C}_{BC}(X \to Y)$	0.0140 (0.0180)	0.0112 (0.0174)	0.0657 (0.0507)	0.0300 (0.0372)	0.0456 (0.0427)	0.0644 (0.0569)	0.0642 (0.0582)	0.0673 (0.0514)	0.0721 (0.0640)	0.0284 (0.2640)	0.1360 (0.1017)
$Y \to X$		No	No	No	No	No	No	No	No	No	No	Yes
True	$C(Y \to X)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1273
Bias corrected	$\hat{C}_{BC}(Y \to X)$	0.0131 (0.0187)	0.0135 (0.0220)	0.0068 (0.0157)	0.0066 (0.0172)	0.0065 (0.0173)	0.0074 (0.0177)	0.0071 (0.0181)	0.0058 (0.0158)	0.0067 (0.0174)	0.0065 (0.0165)	0.1122 (0.0934)
Sample size: $T = 150$												
$X \to Y$		No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
True	$C(X \to Y)$	0.0000	0.0000	0.0771	0.0385	0.0560	0.0738	0.0772	0.0769	0.0786	0.1061	0.1565
Bias corrected	$\hat{C}_{BC}(X \to Y)$	0.0038 (0.0056)	0.0034 (0.0060)	0.0731 (0.0287)	0.0370 (0.0206)	0.0526 (0.0240)	0.0709 (0.0289)	0.0742 (0.0319)	0.0755 (0.0029)	0.0750 (0.0345)	0.0761 (0.1408)	0.1535 (0.6310)
$Y \to X$		No	No	No	No	No	No	No	No	No	No	Yes
True	$C(Y \to X)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1273
Bias corrected	$\hat{C}_{BC}(Y \to X)$	0.0038 (0.0057)	0.0039 (0.0660)	0.0015 (0.0047)	0.0016 (0.0045)	0.0017 (0.0047)	0.0015 (0.0460)	0.0013 (0.0045)	0.0016 (0.0050)	0.0011 (0.0046)	0.0018 (0.0047)	0.1212 (0.5460)

Note: This table shows the values of the bias-corrected estimators of Granger causality. The values in the "True"rows correspond to the theoretical (in the case of Granger non-causality) values, as well as the values of the causality measures that have been obtained using the large simulations algorithm. "No"indicates the absence of a causal relationship, while "Yes"implies the presence of a causal relationship. The upper portion of the table shows the value of the causality measures for a sample size of $T = 50$ and the bottom portion for $T = 150$. The values in the parenthesis are the standard deviation of the estimated values.

Table 3.4: Bootstrap bias-corrected estimators of the sign-based Granger causality measures for $T = 250$ and $T = 500$

	Measure	DGP1	DGP2	DGP3(I)	DGP3(II)	DGP3(III)	DGP3(IV)	DGP3(V)	DGP3(VI)	DGP3(VII)	DGP3(VIII)	DGP4
Sample size: $T = 250$												
$X \to Y$		No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
True	$C(X \to Y)$	0.0000	0.0000	0.0771	0.0385	0.0560	0.0738	0.0772	0.0769	0.0786	0.1061	0.1565
Bias corrected	$\hat{C}_{BC}(X \to Y)$	0.0027	0.0019	0.0761	0.0377	0.0549	0.0720	0.0751	0.0756	0.0780	0.0892	0.1539
		(0.0039)	(0.0032)	(0.0235)	(0.0157)	(0.0171)	(0.0245)	(0.0245)	(0.0210)	(0.0259)	(0.1216)	(0.0511)
$Y \to X$		No	No	No	No	No	No	No	No	No	No	Yes
True	$C(Y \to X)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1273
Bias corrected	$\hat{C}_{BC}(Y \to X)$	0.0029	0.0024	0.0007	0.0006	0.0009	0.0009	0.0012	0.0006	0.0007	0.0010	0.1261
		(0.0041)	(0.0034)	(0.0028)	(0.0025)	(0.0025)	(0.0027)	(0.0029)	(0.0024)	(0.0025)	(0.0029)	(0.0462)
Sample size: $T = 500$												
$X \to Y$		No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
True	$C(X \to Y)$	0.0000	0.0000	0.0771	0.0385	0.0560	0.0738	0.0772	0.0769	0.0786	0.1061	0.1565
Bias corrected	$\hat{C}_{BC}(X \to Y)$	0.0013	0.0010	0.0761	0.0373	0.0556	0.0724	0.0772	0.0778	0.0786	0.0922	0.1542
		(0.0020)	(0.0018)	(0.0159)	(0.0107)	(0.0119)	(0.0161)	(0.0165)	(0.0158)	(0.0182)	(0.0822)	(0.0374)
$Y \to X$		No	No	No	No	No	No	No	No	No	No	Yes
True	$C(Y \to X)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1273
Bias corrected	$\hat{C}_{BC}(Y \to X)$	0.0012	0.0011	0.0004	0.0003	0.0004	0.0003	0.0004	0.0003	0.0004	0.0003	0.1258
		(0.0018)	(0.0018)	(0.0012)	(0.0011)	(0.0012)	(0.0012)	(0.0013)	(0.0012)	(0.0013)	(0.0012)	(0.0335)

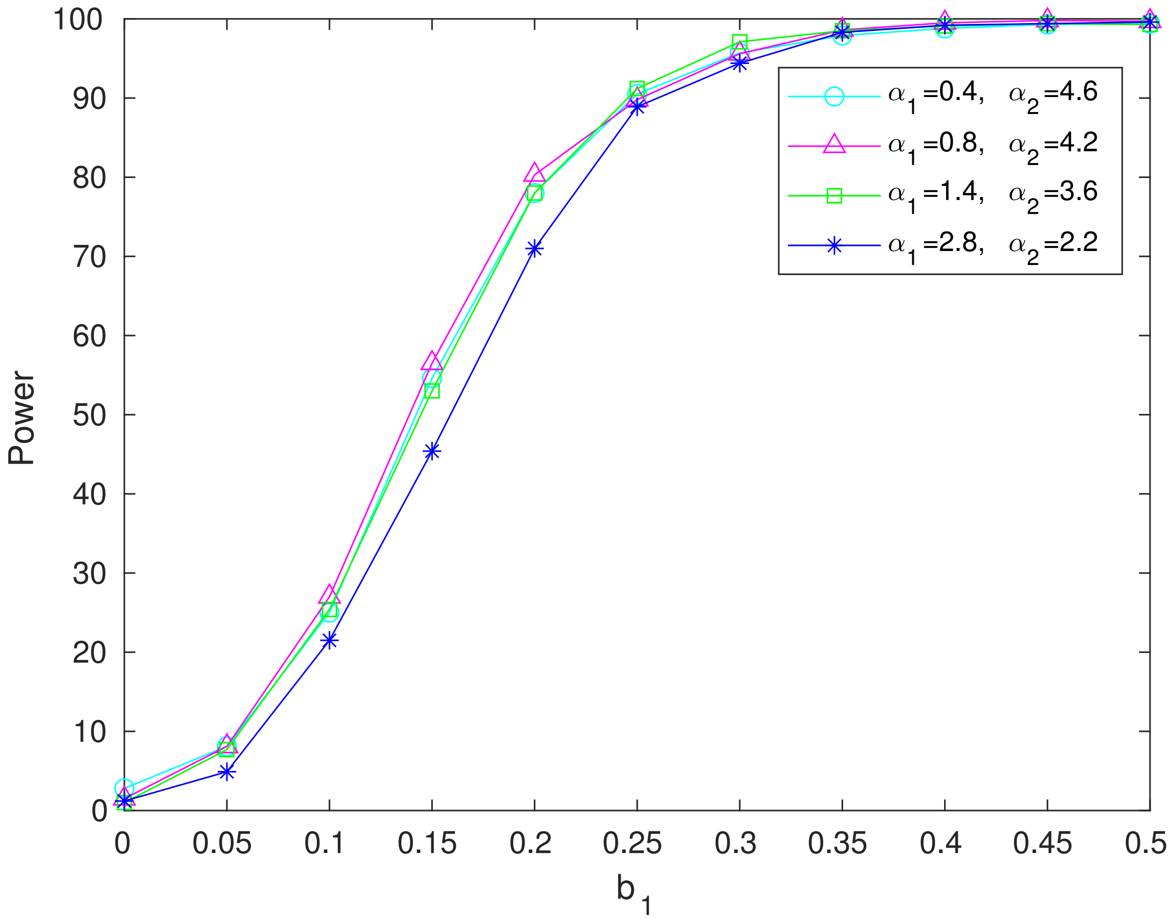
Note: This table shows the values of the bias-corrected estimators of Granger causality. The values in the "True" rows correspond to the theoretical (in the case of Granger non-causality) values, as well as the values of the causality measures that have been obtained using the large simulations algorithm. "No" indicateds the absence of a causal relationship, while "Yes" imlies the presence of a causal relationship. The upper portion of the table shows the value of the causality measures for a sample size of $T = 250$ and the bottom portion for $T = 500$. The values in the paranthesis are the standard deviation of the estimated values.

and α_3 which correspond to the level of the tests based on the elements in $CS_A(\alpha_1)$ are obtained by $\alpha_2 = \alpha - \alpha_1$ and $\alpha_3 = \alpha + \alpha_1$ respectively. In our study, we fix α at 0.05 for a Sample size of $T = 200$, and consider the values of $0.028, 0.014, 0.008$ and 0.004 for α_1. The bounds type procedure entails rejecting the null hypothesis of Granger non-casuality, if $\widetilde{SG}_T(A, \beta_1)$ is significant for each A in $CS_A(\alpha_1)$ at level $\alpha - \alpha_1$, and accepting the null hypothesis if no $\widetilde{SG}_T(A, \beta_1)$ is significant at level $\alpha + \alpha_1$; otherwise, the test is inconclusive. It is evident from figure 3.1 that a wider confidence set for the vector of nuisance parameters $CS_A(\alpha_1)$ leads to a more powerful test, albeit there are diminishing gains in terms of power as the width of the confidence set is increased wider than a threshold -e.g. by choosing α_1 at 0.004. These results are consistent with the findings of Campbell and Dufour (1997). Therefore, in our simulation study, we assign a value of 0.008 to α_1.

Tables 3.6 and 3.7 present the results of the power simulations of the sign-based Granger non-causality tests in turn using the bound-type procedure and the estimation of the nuisance parameter vector approach. The latter involves first estimating the nuisance parameter vector, A, under the null hypothesis of Granger non-causality using the OLS or a robust estimator of choice. Once $\hat{A}$ is obtained, the model is transformed and the test rejects the null hypothesis if $\widetilde{SG}_T(\hat{A}, \beta_1) > c_1(\alpha, \beta_1)$, where $c_1(\alpha, \beta_1)$ is the $\alpha = 5\%$ significance level. First, let us consider DGPs 1 and 2: in both cases there is no causality from X to Y or from Y to X. Evidently, for both samples and both testing procedures the sign-based tests control the size.

Second, for DGPs 3(I)-4(VIII) in which X Granger causes Y but not the converse, we have considered different distributional assumptions for the marginal processes Y and X. In case of the marginal process X, our simulations show that the proposed bound-type tests and those based on the estimation of the nusiance vector, A, control size whatever the sample size. For the marginal process Y and a sample size of $T = 50$, the bound-type procedure possesses good power properties regardless of the distribution; however, DGPs 3(II) and 3(VIII) which correspond to the Cauchy distribution and Normal distribution with GARCH(1,1) variance seem to have the least power among all the different distributional assumptions and forms of heteroskedasticity. Once the sample size is increased to $T = 150$, the power of the test is unity regardless of the distribution.

Figure 3.1: Power simulations for the bound-type procedure with different values of α_1 and α_2



Note: This figure presents the power simulations for the bound-type procedure suggested in Section 3.4.2 for different widths of the confidence set $CS_A(\alpha_1)$. The simulations are considered for testing causality from X to Y for the following VAR(1) process

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} 0.5 & b_1 \\ 0 & 0.9 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} \varepsilon_t^y \\ \varepsilon_t^x \end{bmatrix}}_{\varepsilon_t}, \quad \varepsilon_t \mid y_t, x_t \sim N(0, \Sigma)$$

with a sample size of $T = 200$ and a significance level $\alpha = 0.05$. As it is evident, under all combinations of α_1 and α_2 size is controlled, with the power curve being traced out when a wider confidence set of $\alpha_1 = 0.008$ is considered. However, there appears to be diminishing returns in terms of power as α_1 is reduced to 0.004.

Table 3.5: Power simulations for the bound-type procedure with different values of α_1 and α_2

Testing strategy			Bounds test		
α_1	α_2	b_1	Reject	Accept	Inconclusive
2.8%	2.2%	0.00	1.2	69.4	29.4
		0.05	4.9	48.9	46.2
		0.10	21.5	17.7	60.8
		0.15	45.4	4.5	50.1
		0.20	71.0	0.2	28.8
		0.25	88.9	0.0	11.1
		0.30	94.4	0.0	5.6
		0.35	98.3	0.0	1.7
		0.40	99.2	0.0	0.8
		0.45	99.4	0.0	0.6
		0.50	99.6	0.0	0.4

Testing strategy			Bounds test		
α_1	α_2	b_1	Reject	Accept	Inconclusive
1.4%	3.6%	0.00	1.0	71.9	27.1
		0.05	7.7	53.0	39.3
		0.10	25.4	18.5	56.1
		0.15	53.0	4.5	42.5
		0.20	78.0	0.5	21.5
		0.25	91.2	0.1	8.7
		0.30	97.1	0.1	2.8
		0.35	98.5	0.0	1.5
		0.40	99.1	0.0	0.9
		0.45	99.4	0.0	0.6
		0.50	99.3	0.0	0.7

Testing strategy			Bounds test		
α_1	α_2	b_1	Reject	Accept	Inconclusive
0.8%	4.2%	0.00	1.5	72.7	25.8
		0.05	8.1	54.5	37.5
		0.10	27.0	20.7	52.3
		0.15	56.5	2.4	41.1
		0.20	80.3	0.8	18.9
		0.25	89.8	0.1	10.1
		0.30	95.6	0.0	4.4
		0.35	98.6	0.0	1.4
		0.40	99.5	0.0	0.5
		0.45	99.8	0.0	0.2
		0.50	99.7	0.0	0.3

Testing strategy			Bounds test		
α_1	α_2	b_1	Reject	Accept	Inconclusive
0.4%	4.6%	0.00	2.8	76.2	21.0
		0.05	8.1	55.7	36.2
		0.10	25.0	22.9	52.1
		0.15	54.6	5.3	40.1
		0.20	78.0	0.6	21.4
		0.25	90.5	0.1	9.4
		0.30	95.7	0.0	4.3
		0.35	97.9	0.0	2.1
		0.40	98.8	0.0	1.2
		0.45	99.3	0.0	0.7
		0.50	99.4	0.0	0.6

Note: We consider a sample size of $T = 200$ and a significane level of $\alpha = 0.05$. We consider different testing strategies involving different widths of the confidence set $CS_A(\alpha_1)$ using different combintations of α_1 and α_2. The columns "Reject" indicate that for each A in $CS_A(\alpha_1)$ the test $\widetilde{SG}_T(A, \beta_1)$ is significant at level $\alpha - \alpha_1$, while the columns "Accept" imply that no test is significant at level $\alpha + \alpha_1$; otherwise, the test is inconclusive.

Table 3.6: Power simulations for the bound-type procedure under different distributions and forms of heteroskedasticity.

	DGP1	DGP2	DGP3(I)	DGP3(II)	DGP3(III)	DGP3(IV)	DGP3(V)	DGP3(VI)	DGP3(VII)	DGP3(VIII)	DGP4
Sample size: $T = 50, \quad \alpha = 5\%$											
$X \to Y$											
Reject	1.9	2.2	87.2	88.4	89.1	96.7	94.8	57.7	100	99.9	97.4
Accept	78.5	70.0	0.5	0.8	0.2	0.2	0.0	0.6	0.0	0.0	0.2
$Y \to X$											
Reject	2.3	1.7	0.8	1.4	1.3	1.6	1.7	1.6	1.9	1.3	16.2
Accept	78.9	72.6	70.9	70.3	72.0	70.4	75.0	68.7	71.2	70.4	1.7
Sample size: $T = 100, \quad \alpha = 5\%$											
$X \to Y$											
Reject	3.9	3.0	99.8	99.7	100	100	100	82.2	100	100	100
Accept	80.6	74.6	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0
$Y \to X$											
Reject	4.1	3.3	1.9	3.0	2.1	2.2	2.2	2.7	2.1	2.8	58.6
Accept	79.3	74.2	74.4	70.6	72.8	76.7	78.3	74.2	73.9	71.2	0.0

Note: This table shows the power simulations for the bound-type procedure for sample sizes of $T = 50$ and $T = 150$. The rows "Reject" indicate that for each A in $CS_A(\alpha_1)$ the tests $\widetilde{SG}_T(A, \beta_1)$ is significant at $\alpha - \alpha_1$, while the rows "Accept" that none are significant at $\alpha + \alpha_1$.

Table 3.7: Power simulations with the estimated nuisance parameter vector A under different distributions and forms of heteroskedasticity

	DGP1	DGP2	DGP3(I)	DGP3(II)	DGP3(III)	DGP3(IV)	DGP3(V)	DGP3(VI)	DGP3(VII)	DGP3(VIII)	DGP4
Sample size: $T = 50$, $\alpha = 5\%$											
$X \to Y$											
Reject	4.8	6.1	96.4	77.5	89.2	86.2	94.9	95.4	96.7	52.4	99.1
$Y \to X$											
Reject	4.8	5.8	7.0	5.8	5.0	6.0	6.4	7.5	4.8	5.9	100
Sample size: $T = 150$, $\alpha = 5\%$											
$X \to Y$											
Reject	4.9	5.0	100	100	100	99.7	100	100	100	95	100
$Y \to X$											
Reject	5.5	5.2	5.5	3.9	6.3	5.5	6.0	4.7	5.1	4.8	100

Note: The power simulations in this table consider the methodology that involves the tests that deal with the nusiance parameter problem by first estimating the nuisance parameter vector A under the null hypothesis of Granger non-casuality. Thereafter, the model is transormed and the test is rejected if $\widetilde{SG}_T(\hat{A}, \beta_1) > c_1(\alpha, \beta_1)$.

Finally, in DGP4 we have bidirectional causality, as well as instantaneous causality. Evidently in this case the power of both test is almost at unity regardless of the sample size.

3.6 Empirical application: exchange rates and stock market returns

Numerous studies have investigated the causal linkages between stock market prices and exchange rates with mixed conclusions regarding the direction of causality. Exchange rates are said to have a causal impact on stock market prices, since theoretically speaking, changes in the former affects the firms' profits, which in turn has an impact on the stock market returns [see Aggarwal (1998)]. On the other hand, cases in support of the causal impact of the stock market returns on exchange rates argue that exogenous appreciation of the domestic stock prices increases wealth and consumption, which lead to higher interest rates. This results in higher capital inflows and the appreciation of the domestic currency [See Krueger (1983)]. Therefore, assessing the causal relationship between exchange rates and stock market prices is of great importance for policy makers. If exchanges rates are found to have a causal impact on stock market prices, then government policies on exchange rates may have great ramifications on their stock markets. Furthermore, as suggested by Muhammad et al. (2002), controlling the exchange rates may be used as means of circumventing future stock market crises, or as tools for attracting foreign portfolio investments. If, however, unidirectional causality is found to be stretched from stock prices to exchange rates, policy makers can implement strategies to stabilize the domestic stock markets. Finally, the presence of bidirectional causality could be exploited by investors to make strategic investment decisions, given the information on the other market.

Early studies of the causal linkages between exchange rates and stock market prices include the study by Aggarwal (1998), who found a positive and significant relationship between US stock prices and the trade-weighted US dollar; Soenen and Hennigar (1988), found a negative and significant correlation between the two variables in the monthly data spanning from 1980 to 1986, and Soenen and Aggarwal (1989), obtained mixed results. More recent studies that have utilized more sophisticated econometric tools include the study of Bahmani-Oskooee and Sohrabian (1992), who employed the Granger causality concept to discover a bidirectional relationship between the stock prices of the S&P500 index and the effective exchange rate of the US dollar, and no long-run relationship using cointegration analysis. Abdalla and Murinde (1997), investigated the said relationship in the emerging financial markets in India, Korea, Pakistan and Philippines, using bivariate vector autoregression on monthly data spanning from 1985 to 1994. Their findings suggests unidirectional causality from exchange rates to stock prices. On the other hand, Hatemi-J and Irandoust (2002), by employing the Granger non-causality testing procedure of Toda and Yamamoto (1995) in a vector autoregressive framework, studied the relationship between the Swedish stock market prices and the Swedish Krona, and found evidence of unidirectional causality from stock market prices to exchange rates. In addition, papers with data spanning throughout the financial crisis, have mostly found a bidirectional causal relationship between the two variables: Olugbenga (2012), found a bivariate relationship using the Johansen integration tests for monthly data spanning from 1985 to 2009; Cakan and Ejara (2013) employed monthly data for twelve emerging markets for the period 1990-2013 and found linear and non-linear bi-directional causality; and Zeren and Koç (2016), specifically studied the dynamic relationship between exchange rates and stock market prices during crises and also gathered evidence in favor of two-way causality.

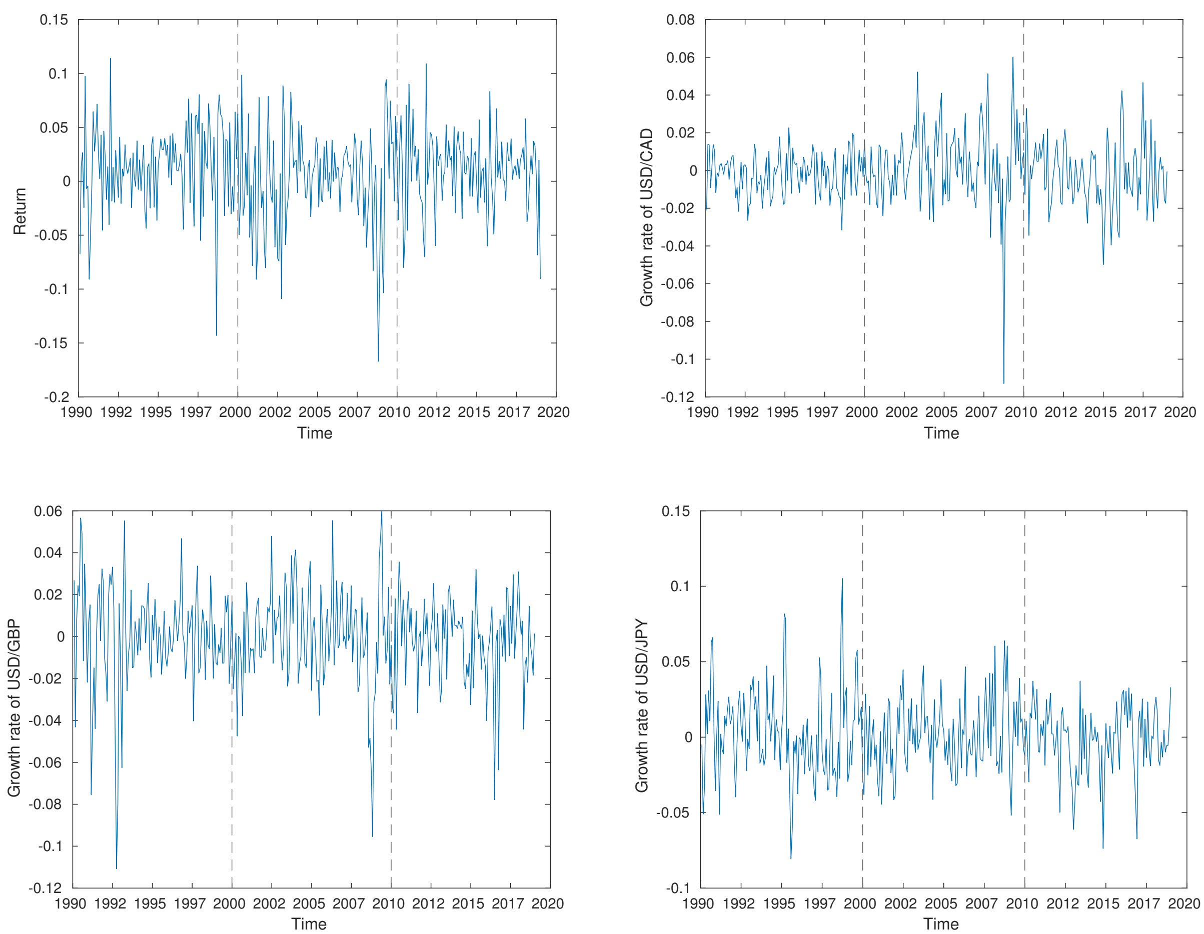
Most studies of the Granger causality relationship between the stock market prices and exchange rates that have been conducted to date may be biased in finite samples due to non-standard distributions and high persistency of the variables. Furthermore, asymptotic approximations are inappropriate in the presence of integrated or cointegrated data [see Hatemi-J and Irandoust (2002)]. Therefore, we utilize the proposed bias-corrected estimator of the sign-based Granger causality measures and the exact sign-based non-causality tests, to assess the strength and significance of the causal relationship between the two variables.

3.6.1 Data description

The monthly prices of the value-weighted S&P500 index and the exchange rates for the US dollars to Canadian dollars, British pounds and the Japanese yen have been retrieved from the Wharton

Research Data Services (WRDS) platform for the period spanning from January 1990 to January 2019 for a total of 348 observations. In order to assess the stationarity of the variables in logarithmic form, we follow the sequential testing strategy of Phillips and Perron (1988) proposed in the first chapter, on both level and first differenced variables using the Augmented Dickey Fuller (ADF) test. Our results suggest that the variables are integrated of the first order, $I(1)$, meaning that they are non-stationary in levels, but stationary when the first differences are calculated. Henceforth, following Taamouti et al. (2014), we will conduct our Granger causality analysis within the framework of returns and growth rates.

Figure 3.2: Monthly S&P500 stock returns and the growth of the USD/CAD, USD/GBP, and USD/JPY exchange rates.



Note: The sample spans from January 1990 to January 2019 for a total of 348 monthly observations. The dashed lines separate the horizon under consideration to three periods: A) January 1990 - January 2000, B) January 2000 - January 2010, and C) January 2010 - January 2019.

3.6.2 Results

We follow Hatemi-J and Irandoust (2002) by combining the Akaike's information criterion (AIC), Bayesian information criterion, and the likelihood ratio (LR) test to choose the order of the VAR(p) model. This strategy involves determining the lag order p by first assessing the outcome of the AIC and BIC. If the results are contradictory, then the LR test is used to choose between the two lag orders. Both criteria using the bivariate VAR(p) process for different combinations of the S&P500 returns and growth of the exchange rates are minimized at a lag order of $p = 1$. Therefore, we consider a VAR(1) model of the form

$$\begin{bmatrix} \Delta \ln P_t \\ \Delta \ln FX_t^i \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} + \begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix} \begin{bmatrix} \Delta \ln P_{t-1} \\ \Delta \ln FX_{t-1}^i \end{bmatrix} + \begin{bmatrix} \varepsilon_t^r \\ \varepsilon_t^{FX} \end{bmatrix}$$

for $i = GBP,\ JPY,\ CAD$. Table 3.8 provides the results of the bias-corrected estimators of the sign-based measures and the exact tests of Granger non-causality for the relationship between the growth in the exchange rates and the stock market returns. We reject the null hypotehsis of Granger non-causality when

$$Q_L(\beta) > c_1(\alpha_2, \beta_1) \tag{3.73}$$

and accept it when

$$Q_U(\beta_1) \leq c_1(a_3, \beta_1) \tag{3.74}$$

otherwise the test is inconclusive. In more simple terms, relationship (3.73) implies that the null hypothesis of Granger non-causality is rejected at the 5% level, if and only if all the tests corresponding to each $A \in CS_A(\alpha_1)$ are rejected at the α_2 level, where $CS_A(\alpha_1)$ is the $1 - \alpha_1$ confidence set for the vector of nuisance parameters that is obtained using the exactly median unbiased estimation procedure of Andrews (1993). Similarly, (3.74) is accepted, if and only if all the tests corresponding to each $A \in CS_A(\alpha_1)$ are accepted at the α_3 level, where α_1, α_2 and α_3 are chosen according to relationship (3.59).

Table 3.8: Results of the causality analysis between the growth of the exchange rates and stock market returns

Direction of causality	Bias-corrected estimate of sign-based measure	$[Q_L(\beta), Q_U(\beta)]$	$[c_1(\alpha_2, \beta_1), c_1(\alpha_3, \beta_1)]$
Panel A: 1990-2000			
USD/CAD $\rightarrow$ r	0.0000	$[0.0425, 0.0548]$	$[0.0473, 0.0457]$
USD/JPY $\rightarrow$ r	0.0000	$[0.0390, 0.0600]$	$[0.0980, 0.0931]$
USD/GBP $\rightarrow$ r	0.0003	$[0.0167, 0.0308]$	$[0.0317, 0.0296]$
r $\rightarrow$ USD/CAD	0.0002	$[0.0893, 0.1022]$	$[0.1064, 0.1038]$
r $\rightarrow$ USD/JPY	0.0005	$[-0.1865, -0.1645]$	$[-0.1152, -0.1222]$
r $\rightarrow$ USD/GBP	0.0000	$[-0.1341, -0.0887]$	$[-0.0689, -0.0764]$
Panel B: 2000-2010			
USD/CAD $\rightarrow$ r	0.0003	$[0.0979, 0.1449]$	$[0.1460, 0.1404]$
USD/JPY $\rightarrow$ r	0.0001	$[-0.0292, 0.0113]$	$[0.0610, 0.0521]$
USD/GBP $\rightarrow$ r	0.0010	$[0.0195, 0.0668]$	$[0.0590, 0.0538]$
r $\rightarrow$ USD/CAD	0.0018	$[\mathbf{0.1172}, \mathbf{0.2346}]$	$[0.1054, 0.0977]$
r $\rightarrow$ USD/JPY	0.0000	$[0.0007, 0.0101]$	$[0.0192, 0.0179]$
r $\rightarrow$ USD/GBP	0.0008	$[0.0621, 0.1175]$	$[0.0741, 0.0685]$
Panel C: 2010-2019			
USD/CAD $\rightarrow$ r	0.0009	$[-0.0666, -0.0056]$	$[0.0134, 0.0042]$
USD/JPY $\rightarrow$ r	0.0000	$[-0.0107, -0.0040]$	$[0.0035, 0.0029]$
USD/GBP $\rightarrow$ r	0.0003	$[0.0511, 0.0739]$	$[0.0758, 0.0723]$
r $\rightarrow$ USD/CAD	0.0000	$[0.0256, 0.0343]$	$[0.0387, 0.0377]$
r $\rightarrow$ USD/JPY	0.0003	$[-0.0481, -0.0215]$	$[-0.0212, -0.0247]$
r $\rightarrow$ USD/GBP	0.0003	$[0.0649, 0.0991]$	$[0.0776, 0.0755]$

Note: This table summarizes the bootstrap bias-corrected estimates of the sign-based Granger causality measure and employs the bound-type sign-based inference procedure to test for Granger non-causality between the variables of growth in exchange rates and the returns on the S&P500 value-weighted index. We reject the null hypothesis, H_0, at the $\alpha = 5\%$ level, when the tests for all $A \in CS_A(\alpha_1)$ are significant at the $\alpha_2 = 0.042$ level $(Q_L(\beta) > c_1(\alpha_2, \beta_1))$ and accept H_0 if none are significant at the $\alpha_3 = 0.058$ level $(Q_U(\beta) \leq c_1(\alpha_3, \beta_1))$. Otherwise the test is "inconclusive". We consider $B = 500$ sample estimates of the causality measure based on the VAR sieve bootstrapped data to calculate the bias term, $\widehat{Bias}_{boot}$. Furthermore, we set α_1 at 0.008 for the confidence set $A \in CS_A(\alpha_1)$, to maximize power. The brackets in bold imply the test is significance of the test at the 5% level, while the rest are either inconclusive or accept H_0.

Panels A and C, which correspond to the periods (January 1990 - January 2000) and (January 2010 - January 2019) respectively, suggest lack of any evidence in favor of a causal relationship between the growths of the exchange rates and the returns on the S&P500 index. The bias-corrected estimates of the sign-based Granger causality measures reveal a relatively weak degree of causality between the said variables, and the sign-based tests fail to reject the null hypothesis of Granger non-causality, either by accepting the null hypothesis or by providing inconclusive results. These findings also generally hold true for the period spanning from January 2000 to January 2010 in panel C. However, the bias-corrected estimate the of sign-based measures is considerably (almost ten-fold) greater than other periods for the returns on the S&P500 index to the USD/CAD exchange rates. Furthermore, the sign-based Granger non-causality test shows significance at the 5% level. However, there is no reverse causality in the same period for these two variables, which implies only the presence of unidirectional causality.

3.7 Conclusion

In this chapter, we propose sign-based Granger causality measures based on the Kullback-Leibler distance to assess the strength of the causal relationship between random variables. These measures are distribution-free and are particularly attractive in case of asymmetric distributions, in which there is no evidence of causality in the the mean. Thereafter, we show that a VAR(p) model can be fitted to the data and propose a consistent estimator for the sign-based measure. In finite samples, we suggested the use of the vector autoregressive sieve bootstrap to calculate the finite sample bias and the bootstrap bias-corrected estimate of the causality measures. In a simulation study, we considered different DGPs that are commonly encountered in practice. The bootstrap bias-corrected estimator of the causality measure performs well and provides evidence in favor of the desired outcome. We then showed that by using the bound-type procedures as in Dufour (1990) and Campbell and Dufour (1997) to address the nuisance parameter problem under the null hypothesis of Granger non-causality, tests of Granger non-causality can be developed as byproduct of the sign-based causality measures. The tests are exact, distribution-free and robust against heteroskedasticity. In a simulation study our procedures are shown to be valid (control size whatever the sample size), robust against heteroskedasticity and possess good power properties. Finally, using the monthly data on the growth in the USD/CAD, USD/GBP and USD/JPY exchange rates and the returns on the S&P500 index, we consider an empirical application to assess the causal linkages between the stock market returns and the growth in the exchange rates. Our results indicate the presence of a unidirectional causality for the returns on the S&P500 index to the USD/CAD exchange rates, as the bias-corrected estimate the of sign-based measures is considerably greater than other periods. Furthermore, the sign-based Granger non-causality test reveal significance at the 5% level.

3.8 Appendix

Proof of Corollary 5. For the unrestricted likelihood function, we have

$$\begin{aligned}
\log\left\{P_\theta[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X]\right\} &= S_t^y \log\left\{\frac{P_\theta[y_t \geq 0 \mid \mathbf{S}_{t-1}^y, Y, X]}{P_\theta[y_t < 0 \mid \mathbf{S}_{t-1}^y, Y, X]}\right\} + \log P_\theta[y_t < 0 \mid \mathbf{S}_{t-1}^y, Y, X] \\
&= S_t^y \left\{\log P_\theta[y_t \geq 0 \mid \mathbf{S}_{t-1}^y, Y, X] - \log P_\theta[y_t < 0 \mid \mathbf{S}_{t-1}^y, Y, X]\right\} \\
&\quad + \log P_\theta[y_t < 0 \mid \mathbf{S}_{t-1}^y, Y, X] \\
&= S_t^y \left\{\begin{array}{l}
\log\Big\{P_\theta[y_t \geq 0 \mid y_{t-1} \geq 0, Y, X]^{S_{t-1}^y} \times \\
\qquad P_\theta[y_t \geq 0 \mid y_{t-1} < 0, Y, X]^{1-S_{t-1}^y}\Big\} - \\
\log\Big\{P_\theta[y_t < 0 \mid y_{t-1} \geq 0, Y, X]^{S_{t-1}^y} \times \\
\qquad P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X]^{1-S_{t-1}^y}\Big\}
\end{array}\right\} \\
&\quad + \log P_\theta[y_t < 0 \mid y_{t-1} \geq 0, Y, X]^{S_{t-1}^y} P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X]^{1-S_{t-1}^y}
\end{aligned}$$

which may further get extended to

$$\begin{aligned}
\log\left\{P_\theta[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X]\right\} &= S_t^y \left\{\begin{array}{l}
S_{t-1}^y \log P_\theta[y_t \geq 0 \mid y_{t-1} \geq 0, Y, X] + \\
(1 - S_{t-1}^y) \log P_\theta[y_t \geq 0 \mid y_{t-1} < 0, Y, X] - \\
S_{t-1}^y \log P_\theta[y_t < 0 \mid y_{t-1} \geq 0, Y, X] - \\
(1 - S_{t-1}^y) \log P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X]
\end{array}\right\} \\
&\quad + S_{t-1}^y \log P_\theta[y_t < 0 \mid y_{t-1} \geq 0, Y, X] \\
&\quad + (1 - S_{t-1}^y) \log P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X] \\
&= S_t^y \left\{\begin{array}{l}
S_{t-1}^y \left\{\log\left\{\frac{P_\theta[y_t \geq 0 \mid y_{t-1} \geq 0, Y, X]}{P_\theta[y_t < 0 \mid y_{t-1} \geq 0, Y, X]}\right\} - \log\left\{\frac{P_\theta[y_t \geq 0 \mid y_{t-1} < 0, Y, X]}{P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X]}\right\}\right\} \\
+ \log\left\{\frac{P_\theta[y_t \geq 0 \mid y_{t-1} < 0, Y, X]}{P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X]}\right\}
\end{array}\right\} \\
&\quad + S_{t-1}^y \log\left\{\frac{P_\theta[y_t < 0 \mid y_{t-1} \geq 0, Y, X]}{P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X]}\right\} + \log P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X]
\end{aligned}$$

$$\begin{aligned}
\log\left\{P_\theta[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X]\right\} &= S_t^y S_{t-1}^y \left\{\begin{array}{l}
\log\left\{\frac{P_\theta[y_t \geq 0 \mid y_{t-1} \geq 0, Y, X]}{P_\theta[y_t < 0 \mid y_{t-1} \geq 0, Y, X]}\right\} - \\
\log\left\{\frac{P_\theta[y_t \geq 0 \mid y_{t-1} < 0, Y, X]}{P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X]}\right\}
\end{array}\right\} \\
&\quad + S_t^y \log\left\{\frac{P_\theta[y_t \geq 0 \mid y_{t-1} < 0, Y, X]}{P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X]}\right\} \\
&\quad + S_{t-1}^y \log\left\{\frac{P_\theta[y_t < 0 \mid y_{t-1} \geq 0, Y, X]}{P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X]}\right\} + \log P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X]
\end{aligned}$$

for $t = 2, ..., T$ we have

$$\begin{aligned}
P_\theta[y_t \geq 0 \mid y_{t-1} < 0, Y, X] &= 1 - P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X] \\
&= 1 - \frac{P_\theta[y_t < 0, y_{t-1} < 0 \mid Y, X]}{P_\theta[y_{t-1} < 0 \mid Y, X]} \\
&= 1 - \frac{P[\varepsilon_t < -\theta' J_{t-1}, \varepsilon_{t-1} < -\theta' J_{t-2} \mid Y, X]}{P[\varepsilon_{t-1} < -\theta' J_{t-2} \mid Y, X]} \\
P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X] &= \frac{P_\theta[y_t < 0, y_{t-1} < 0 \mid Y, X]}{P_\theta[y_{t-1} < 0 \mid Y, X]} \\
&= \frac{P[\varepsilon_t < -\theta' J_{t-1}, \varepsilon_{t-1} < -\theta' J_{t-2} \mid Y, X]}{P[\varepsilon_{t-1} < -\theta' J_{t-2} \mid Y, X]} \\
P_\theta[y_t \geq 0 \mid y_{t-1} \geq 0, Y, X] &= 1 - P_\theta[y_t < 0 \mid y_{t-1} \geq 0, Y, X] \\
&= 1 - \left[\frac{P_\theta[y_t < 0 \mid Y, X]}{P_\theta[y_{t-1} \geq 0 \mid Y, X]} P_\theta[y_{t-1} \geq 0 \mid y_t < 0, Y, X]\right] \\
&= 1 - \left[\frac{P_\theta[y_t < 0 \mid Y, X]}{P_\theta[y_{t-1} \geq 0 \mid Y, X]} (1 - P_\theta[y_{t-1} < 0 \mid y_t < 0, Y, X])\right] \\
&= 1 - \left(\frac{P_\theta[y_t < 0 \mid Y, X]}{P_\theta[y_{t-1} \geq 0 \mid Y, X]} - \frac{P_\theta[y_{t-1} < 0, y_t < 0 \mid Y, X]}{P_\theta[y_{t-1} \geq 0 \mid Y, X]}\right) \\
&= 1 - \left(\frac{P_\theta[y_t < 0 \mid Y, X]}{1 - P_\theta[y_{t-1} < 0 \mid Y, X]} - \frac{P_\theta[y_{t-1} < 0, y_t < 0 \mid Y, X]}{1 - P_\theta[y_{t-1} < 0 \mid Y, X]}\right) \\
&= 1 - \left(\frac{P[\varepsilon_t < -\theta' J_{t-1} \mid Y, X]}{1 - P[\varepsilon_{t-1} < -\theta' J_{t-2} \mid Y, X]}\right. \\
&\qquad \left. - \frac{P[\varepsilon_{t-1} < -\theta' J_{t-2}, \varepsilon_t < -\theta' J_{t-1} \mid Y, X]}{1 - P[\varepsilon_{t-1} < -\theta' J_{t-2} \mid Y, X]}\right) \\
P_\theta[y_t < 0 \mid y_{t-1} \geq 0, Y, X] &= \frac{P_\theta[y_t < 0 \mid Y, X]}{P_\theta[y_{t-1} \geq 0 \mid Y, X]} P_\theta[y_{t-1} \geq 0 \mid y_t < 0, Y, X] \\
&= \frac{P_\theta[y_t < 0 \mid Y, X]}{P_\theta[y_{t-1} \geq 0 \mid Y, X]} (1 - P_\theta[y_{t-1} < 0 \mid y_t < 0, Y, X]) \\
&= \frac{P_\theta[y_t < 0 \mid Y, X]}{P_\theta[y_{t-1} \geq 0 \mid Y, X]} - \frac{P_\theta[y_{t-1} < 0, y_t < 0 \mid Y, X]}{P_\theta[y_{t-1} \geq 0 \mid Y, X]} \\
&= \frac{P_\theta[y_t < 0 \mid Y, X]}{1 - P_\theta[y_t < 0 \mid Y, X]} - \frac{P_\theta[y_{t-1} < 0, y_t < 0 \mid Y, X]}{1 - P_\theta[y_t < 0 \mid Y, X]} \\
&= \frac{P[\varepsilon_t < -\theta' J_{t-1} \mid Y, X]}{1 - P[\varepsilon_t < -\theta' J_{t-1} \mid Y, X]} - \frac{P[\varepsilon_{t-1} < -\theta' J_{t-2}, \varepsilon_t < -\theta' J_{t-1} \mid Y, X]}{1 - P[\varepsilon_t < -\theta' J_{t-1} \mid Y, X]}
\end{aligned}$$

Similarly for the restricted case we have

$$
\begin{aligned}
P_{\theta^R}[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y] \quad = \quad & S_t^y S_{t-1}^y \left\{ \log \left\{ \frac{P_{\theta^R}[y_t \geq 0 \mid y_{t-1} \geq 0, Y]}{P_{\theta^R}[y_t < 0 \mid y_{t-1} \geq 0, Y]} \right\} - \log \left\{ \frac{P_{\theta^R}[y_t \geq 0 \mid y_{t-1} < 0, Y]}{P_{\theta^R}[y_t < 0 \mid y_{t-1} < 0, Y]} \right\} \right\} \\
& + S_t^y \log \left\{ \frac{P_{\theta^R}[y_t \geq 0 \mid y_{t-1} < 0, Y]}{P_{\theta^R}[y_t < 0 \mid y_{t-1} < 0, Y]} \right\} + S_{t-1}^y \log \left\{ \frac{P_{\theta^R}[y_t < 0 \mid y_{t-1} \geq 0, Y]}{P_{\theta^R}[y_t < 0 \mid y_{t-1} < 0, Y]} \right\} \\
& + \log P_{\theta^R}[y_t < 0 \mid y_{t-1} < 0, Y]
\end{aligned}
$$

where for $t = 2, ..., T$

$$
\begin{aligned}
P_{\theta^R}[y_t \geq 0 \mid y_{t-1} < 0, Y] \quad &= \quad 1 - \frac{P[\varepsilon_t < -\theta'^R J_{t-1}, \varepsilon_{t-1} < -\theta'^R J_{t-2} \mid Y]}{P[\varepsilon_{t-1} < -\theta'^R J_{t-2} \mid Y]} \\
P_{\theta^R}[y_t < 0 \mid y_{t-1} < 0, Y] \quad &= \quad \frac{P[\varepsilon_t < -\theta'^R J_{t-1}, \varepsilon_{t-1} < -\theta'^R J_{t-2} \mid Y]}{P[\varepsilon_{t-1} < -\theta'^R J_{t-2} \mid Y]} \\
P_{\theta^R}[y_t \geq 0 \mid y_{t-1} \geq 0, Y] \quad &= \quad 1 - \left(\frac{P[\varepsilon_t < -\theta'^R J_{t-1} \mid Y]}{1 - P[\varepsilon_{t-1} < -\theta'^R J_{t-2} \mid Y]} - \frac{P[\varepsilon_{t-1} < -\theta'^R J_{t-2}, \varepsilon_t < -\theta'^R J_{t-1} \mid Y]}{1 - P[\varepsilon_{t-1} < -\theta'^R J_{t-2} \mid Y]} \right) \\
P_{\theta^R}[y_t < 0 \mid y_{t-1} \geq 0, Y] \quad &= \quad \frac{P[\varepsilon_t < -\theta'^R J_{t-1} \mid Y]}{1 - P[\varepsilon_{t-1} < -\theta'^R J_{t-2} \mid Y]} - \frac{P[\varepsilon_{t-1} < -\theta'^R J_{t-2}, \varepsilon_t < -\theta'^R J_{t-1} \mid Y]}{1 - P[\varepsilon_{t-1} < -\theta'^R J_{t-2} \mid Y]}
\end{aligned}
$$

therefore

$$
\begin{aligned}
\log \left(\frac{P_\theta[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y, X]}{P_{\theta^R}[S_t^y = s_t^y \mid \mathbf{S}_{t-1}^y, Y]} \right) \quad = \quad & S_t^y S_{t-1}^y \left\{ \begin{array}{c} \log \left\{ \frac{P_\theta[y_t \geq 0 | y_{t-1} \geq 0, Y, X]}{P_\theta[y_t < 0 | y_{t-1} \geq 0, Y, X]} \right\} - \\ \log \left\{ \frac{P_\theta[y_t \geq 0 | y_{t-1} < 0, Y, X]}{P_\theta[y_t < 0 | y_{t-1} < 0, Y, X]} \right\} \end{array} \right\} \\
& + S_t^y \log \left\{ \frac{P_\theta[y_t \geq 0 \mid y_{t-1} < 0, Y, X]}{P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X]} \right\} \\
& + S_{t-1}^y \log \left\{ \frac{P_\theta[y_t < 0 \mid y_{t-1} \geq 0, Y, X]}{P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X]} \right\} + \log P_\theta[y_t < 0 \mid y_{t-1} < 0, Y, X] \\
& - S_t^y S_{t-1}^y \left\{ \begin{array}{c} \log \left\{ \frac{P_{\theta^R}[y_t \geq 0 | y_{t-1} \geq 0, Y]}{P_{\theta^R}[y_t < 0 | y_{t-1} \geq 0, Y]} \right\} - \\ \log \left\{ \frac{P_{\theta^R}[y_t \geq 0 | y_{t-1} < 0, Y]}{P_{\theta^R}[y_t < 0 | y_{t-1} < 0, Y]} \right\} \end{array} \right\} \\
& - S_t^y \log \left\{ \frac{P_{\theta^R}[y_t \geq 0 \mid y_{t-1} < 0, Y]}{P_{\theta^R}[y_t < 0 \mid y_{t-1} < 0, Y]} \right\} \\
& - S_{t-1}^y \log \left\{ \frac{P_{\theta^R}[y_t < 0 \mid y_{t-1} \geq 0, Y]}{P_{\theta^R}[y_t < 0 \mid y_{t-1} < 0, Y]} \right\} - \log P_{\theta^R}[y_t < 0 \mid y_{t-1} < 0, Y]
\end{aligned}
$$

$$
\begin{aligned}
\log\left(\frac{P_\theta[S^y_t = s^y_t \mid \mathbf{S}^y_{t-1}, Y, X]}{P_{\theta^R}[S^y_t = s^y_t \mid \mathbf{S}^y_{t-1}, Y]}\right) \;&=\; S^y_t S^y_{t-1}\left[\begin{array}{l}\left(\log\left\{\frac{P_\theta[y_t\geq 0|y_{t-1}\geq 0,Y,X]}{P_\theta[y_t<0|y_{t-1}\geq 0,Y,X]}\right\} - \log\left\{\frac{P_\theta[y_t\geq 0|y_{t-1}<0,Y,X]}{P_\theta[y_t<0|y_{t-1}<0,Y,X]}\right\}\right) - \\ \left(\log\left\{\frac{P_{\theta^R}[y_t\geq 0|y_{t-1}\geq 0,Y]}{P_{\theta^R}[y_t<0|y_{t-1}\geq 0,Y]}\right\} - \log\left\{\frac{P_{\theta^R}[y_t\geq 0|y_{t-1}<0,Y]}{P_{\theta^R}[y_t<0|y_{t-1}<0,Y]}\right\}\right)\end{array}\right] \\
&\quad + S^y_t\left[\begin{array}{l}\log\left\{\frac{P_\theta[y_t\geq 0|y_{t-1}<0,Y,X]}{P_\theta[y_t<0|y_{t-1}<0,Y,X]}\right\} - \\ \log\left\{\frac{P_{\theta^R}[y_t\geq 0|y_{t-1}<0,Y]}{P_{\theta^R}[y_t<0|y_{t-1}<0,Y]}\right\}\end{array}\right] \\
&\quad + S^y_{t-1}\left[\begin{array}{l}\log\left\{\frac{P_\theta[y_t<0|y_{t-1}\geq 0,Y,X]}{P_\theta[y_t<0|y_{t-1}<0,Y,X]}\right\} - \\ \log\left\{\frac{P_{\theta^R}[y_t<0|y_{t-1}\geq 0,Y]}{P_{\theta^R}[y_t<0|y_{t-1}<0,Y]}\right\}\end{array}\right] + \log\left\{\frac{P_\theta[y_t<0 \mid y_{t-1}<0,Y,X]}{P_{\theta^R}[y_t<0 \mid y_{t-1}<0,Y]}\right\}
\end{aligned}
$$

which may further expand to

$$
\begin{aligned}
\log\left(\frac{P_\theta[S^y_t = s^y_t \mid \mathbf{S}^y_{t-1}, Y, X]}{P_{\theta^R}[S^y_t = s^y_t \mid \mathbf{S}^y_{t-1}, Y]}\right) &= \log\big(P_\theta[S^y_t = s^y_t \mid \mathbf{S}^y_{t-1}, Y, X]\big) - \log\big(P_{\theta^R}[S^y_t = s^y_t \mid \mathbf{S}^y_{t-1}, Y]\big) \\
&= S^y_t S^y_{t-1}\alpha_t(\theta/\theta^R) + S^y_t\beta_t(\theta/\theta^R) \\
&\quad + S^y_{t-1}\gamma_t(\theta/\theta^R) + \delta_t(\theta/\theta^R)
\end{aligned}
$$

where for $t = 2, ..., T$

$$
\begin{aligned}
\alpha_t(\theta/\theta^R) \;=\; &\Bigg[\Bigg(\log\left\{\frac{1-\left(\frac{P[\varepsilon_t<-\theta' J_{t-1}|Y,X]}{1-P[\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]} - \frac{P[\varepsilon_t<-\theta' J_{t-1},\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}{1-P[\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}\right)}{\frac{P[\varepsilon_t<-\theta' J_{t-1}|Y,X]}{1-P[\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]} - \frac{P[\varepsilon_{t-1}<-\theta' J_{t-2},\varepsilon_t<-\theta' J_{t-1}|Y,X]}{1-P[\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}}\right\} \\
&- \log\left\{\frac{1-\frac{P[\varepsilon_t<-\theta' J_{t-1},\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}{P[\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}}{\frac{P[\varepsilon_t<-\theta' J_{t-1},\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}{P[\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}}\right\}\Bigg) \\
&- \Bigg(\log\left\{\frac{1-\left(\frac{P[\varepsilon_t<-\theta'^R J_{t-1}|Y]}{1-P[\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]} - \frac{P[\varepsilon_{t-1}<-\theta'^R J_{t-2},\varepsilon_t<-\theta'^R J_{t-1}|Y]}{1-P[\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}\right)}{\frac{P[\varepsilon_t<-\theta'^R J_{t-1}|Y]}{1-P[\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]} - \frac{P[\varepsilon_{t-1}<-\theta'^R J_{t-2},\varepsilon_t<-\theta'^R J_{t-1}|Y]}{1-P[\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}}\right\} \\
&- \log\left\{\frac{1-\frac{P[\varepsilon_t<-\theta'^R J_{t-1},\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}{P[\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}}{\frac{P[\varepsilon_t<-\theta'^R J_{t-1},\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}{P[\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}}\right\}\Bigg)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\beta_t(\theta/\theta^R) &= \left[\log\left\{\frac{1-\frac{P[\varepsilon_t<-\theta' J_{t-1},\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}{P[\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}}{\frac{P[\varepsilon_t<-\theta' J_{t-1},\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}{P[\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}}\right\}\right.\\
&\quad\left.-\log\left\{\frac{1-\frac{P[\varepsilon_t<-\theta'^R J_{t-1},\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}{P[\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}}{\frac{P[\varepsilon_t<-\theta'^R J_{t-1},\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}{P[\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}}\right\}\right]\\
\gamma_t(\theta/\theta^R) &= \left[\log\left\{\frac{\frac{P[\varepsilon_t<-\theta' J_{t-1}|Y,X]}{1-P[\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}-\frac{P[\varepsilon_{t-1}<-\theta' J_{t-2},\varepsilon_t<-\theta' J_{t-1}|Y,X]}{1-P[\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}}{\frac{P[\varepsilon_t<-\theta' J_{t-1},\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}{P[\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}}\right\}\right.\\
&\quad\left.-\log\left\{\frac{\frac{P[\varepsilon_t<-\theta'^R J_{t-1}|Y]}{1-P[\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}-\frac{P[\varepsilon_{t-1}<-\theta'^R J_{t-2},\varepsilon_t<-\theta'^R J_{t-1}|Y]}{1-P[\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}}{\frac{P[\varepsilon_t<-\theta'^R J_{t-1},\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}{P[\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}}\right\}\right]\\
\delta_t(\theta/\theta^R) &= \log\left\{\frac{\frac{P[\varepsilon_t<-\theta' J_{t-1},\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}{P[\varepsilon_{t-1}<-\theta' J_{t-2}|Y,X]}}{\frac{P[\varepsilon_t<-\theta'^R J_{t-1},\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}{P[\varepsilon_{t-1}<-\theta'^R J_{t-2}|Y]}}\right\}
\end{aligned}
$$

and

$$
\begin{aligned}
\alpha_1(\theta/\theta^R) &= 0,\\
\beta_1(\theta/\theta^R) &= \left[\log\left\{\frac{1-P[\varepsilon_1<-\theta' J_0\mid Y,X]}{P[\varepsilon_1<-\theta' J_0\mid Y,X]}\right\}\right.\\
&\quad\left.-\log\left\{\frac{1-P[\varepsilon_1<-\theta'^R J_0\mid Y]}{P[\varepsilon_1<-\theta'^R J_0\mid Y]}\right\}\right]\\
&= \log\left\{\frac{[1-P[\varepsilon_1<-\theta' J_0\mid Y,X]P[\varepsilon_1<-\theta'^R J_0\mid Y]}{(1-P[\varepsilon_1<-\theta'^R J_0\mid Y])P[\varepsilon_1<-\theta' J_0\mid Y,X]}\right\}\\
\gamma_1(\theta/\theta^R) &= 0,\\
\delta_1(\theta/\theta^R) &= \log\left\{\frac{P[\varepsilon_1<-\theta' J_0\mid Y,X]}{P[\varepsilon_1<-\theta'^R J_0\mid Y]}\right\}
\end{aligned}
$$

■

Proof of Proposition 5. Consider the stochastic process $\{J_t = (y_t, x_t) : \Omega \to \mathbb{R}\times\mathbb{R} \equiv \mathbb{R}^2\}_{t=0,1,2,...}$ defined on the probability space $(\Omega, \mathcal{F}, P)$. Let us denote

$$q_t(J_{t-1}, \hat{\theta}, \hat{\theta}^R) = \left\{S_t^y S_{t-1}^y \alpha_t(\hat{\theta}/\hat{\theta}^R) + S_t^y \beta_t(\hat{\theta}/\hat{\theta}^R) + S_{t-1}^y \gamma_t(\hat{\theta}/\hat{\theta}^R) + \delta_t(\hat{\theta}/\hat{\theta}^R)\right\}, \quad t = 1, ..., T.$$

Assumption A4 (1) allows the data to possess temporal dependence and be heterogeneously distributed. Furthermore, as a consequence of this assumption, $\{J_{t-1}\varepsilon_t^y\}$ and $\{J_{t-1}J_{t-1}'\}$ are also α-mixing sequences of size $-r/r-1$ for $r > 1$ [see Proposition 3.50 of White (2014)]. Finally, given the moment conditions A4 (2ii)-(2iii), $\mathbf{J}'\varepsilon^y/n \xrightarrow{a.s.} 0$ and $\mathbf{J}'\mathbf{J}/n - M_n \xrightarrow{a.s.} 0$, where $M_n = O(1)$ and is uniformly positive definite as per assumption A4 (4). Therefore, given Theorem 2.18 of White (2014)

$$\hat{\theta} \xrightarrow{a.s.} \theta, \tag{3.75}$$

and similarly

$$\hat{\theta}^R \xrightarrow{a.s.} \theta^R, \tag{3.76}$$

as a result of which

$$q_t(J_{t-1}, \hat{\theta}, \hat{\theta}^R) \xrightarrow{a.s.} q_t(J_{t-1}, \theta, \theta^R), \quad t = 1, ..., T. \tag{3.77}$$

To prove the consistency of the estimator of the sign-based measures, we follow Coudin and Dufour (2004) by first showing **pointwise convergence**, followed by proving the **uniform convergence** of the estimator. To achieve the former, we must show that $T^{-1} \sum_{t=1}^{T} q_t(J_{t-1}, \theta, \theta^R) - \mathbb{E}\left[q_t(J_{t-1}, \theta, \theta^R)\right]$ converges in probability to zero for all $\theta, \theta^R \in \Theta$. The mixing assumption A4 (1) for J is exported to $q_t(J_{t-1}, \theta, \theta^R)$ meaning that $q_t(J_{t-1}, \theta, \theta^R)_{t=1,2,...}$ is a mixing sequence of α of size $-r/(r-1)$, $r > 1$. Hence, in conjunction with assumption A4 (2i) and in accordance to Corollary 3.48 of White (2014)

$$\frac{1}{T} \sum_{t=1}^{T} q_t(J_{t-1}, \theta.\theta^R) - \mathbb{E}\left[q_t(J_{t-1}, \theta, \theta^R)\right] \xrightarrow[T\to\infty]{p} 0, \quad \forall \theta, \theta^R \in \Theta. \tag{3.78}$$

Next step consists of proving uniform convergence - i.e. we wish to show that $\sup_{\theta,\theta^R\in\Theta} \left| T^{-1} \sum_{t=1}^{T} q_t(J_{t-1}, \theta, \theta^R) - \mathbb{E}\left[q_t(J_{t-1}, \theta, \theta^R)\right] \right|$ converges in probability to zero. To accomplish this, following Andrews (1987): let $B(\tilde{\theta}, \rho)$ be an open ball around $\tilde{\theta}$ with radius ρ, where $\tilde{\theta} = (\theta, \theta^R)$ - in other words, $B(\tilde{\theta}, \rho) = \{\tilde{\theta}' \in \Theta : d(\tilde{\theta}', \tilde{\theta}) < \rho\}$, where $\tilde{\theta}' = (\theta', \theta'^R)$ and $d(.)$ is a metric on Θ. Furthermore, we define

$$\begin{aligned} q_t^+(J_{t-1}, \tilde{\theta}, \rho) &= \sup_{\tilde{\theta}' \in B(\tilde{\theta}, \rho)} q_t(J_{t-1}, \tilde{\theta}'), \\ q_t^-(J_{t-1}, \tilde{\theta}, \rho) &= \inf_{\tilde{\theta}' \in B(\tilde{\theta}, \rho)} q_t(J_{t-1}, \tilde{\theta}'). \end{aligned}$$

Assumptions A1 and B1 of Andrews (1987) correspond to the compactness assumption A4 (5) and mixing assumption A4 (1) respectively. Assumption B2 of Andrews (1987) on the other hand, first requires $q_t^+(J_{t-1}, \tilde{\theta}, \rho)$, $q_t^-(J_{t-1}, \tilde{\theta}, \rho)$ and $q_t(J_{t-1})$ to be random variables and $q_t^+(., \tilde{\theta}, \rho)$ and $q_t^-(., \tilde{\theta}, \rho)$ to be measurable functions from $(\Omega, \mathcal{F}, P)$ to $(\mathbb{R}, \mathcal{B})$, $\forall t$, $\tilde{\theta} \in \Theta$ and ρ, where $\mathcal{B}$ is Borel σ-algebra on $\mathbb{R}$. Furthermore, this assumption requires that $\sup_t \mathbb{E} q_t(J_{t-1})^{r+\delta} < \infty$ for some $\delta > 0$. The mixing condition A4 (1) ensures measurability, and the boundedness condition A4 (3) satisfies the second requirement of this assumption.

Finally, let μ be a σ-finite measure that dominates the marginal distributions of J_{t-1}, $t = 1, 2, ...$ and denote $p_t(j)$ as the density of J_{t-1} w.r.t μ. Assumption A6 of Andrews (1987) then requires that $q_t(J_{t-1}, \tilde{\theta}) p_t(j)$ is continuous in $\tilde{\theta} = \tilde{\theta}^*$ uniformly in t a.e. w.r.t. μ, for each $\tilde{\theta}^* \in \Theta$, and $q_t(J_{t-1}, \tilde{\theta})$ is measerable w.r.t. to the Borel measure for each t and each $\tilde{\theta} \in \Theta$ and

$$\int \sup_{t \geq 1, \tilde{\theta} \in \Theta} |q_t(J_{t-1}, \tilde{\theta})| p_t(j) d\mu(j) < \infty.$$

Note that assumptions (3.22) implies that the residuals ε_t^y have no mass at zero, i.e. $P[\varepsilon_t^y = 0 \mid Y, X] = 0 \quad \forall t$, an assumption that is satisfied when ε_t^y is continuous. Thus, $P[\varepsilon_t^y = -\theta' J_{t-1} \mid Y, X] = 0$, $\forall \theta$ uniformly in t, satisfying the requirement that $q_t(J_{t-1}, \tilde{\theta})$ is continuous in $\tilde{\theta}$ everywhere. Finally, $q_t(J_{t-1}, \tilde{\theta})$ is L_1 bounded and uniformly integrable, satisfying condition A6 of Andrews (1987). Hence, according to the main Theorem and Corollaries 1 and 3 of Andrews (1987), since assumptions A1, B1, B2 and A6 hold

a) $\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[q_t(J_{t-1}, \tilde{\theta})\right]$ is continuous on Θ uniformly over $n \geq 1$,

b) $\sup_{\tilde{\theta} \in \Theta} \left| n^{-1} \sum_{t=1}^{n} q_t(J_{t-1}, \tilde{\theta}) - \mathbb{E}\left[q_t(J_{t-1}, \tilde{\theta})\right] \right| \xrightarrow[n \to \infty]{p} 0$, where $n = (T - p) + 1$.

■

Exact confidence interval algorithm. To calculate the quantiles, $q_{p_j}(\hat{a}_1)$ of $\hat{a}_1$ for $j = 1, 2$, Andrews (1993) proposes a number of approaches. The objective of each approach is to find the value of c that corresponds to

$$P[\hat{a}_1 \leq c] = q_{p_j}(\hat{a}_1), \quad j = 1, 2. \tag{3.79}$$

We know that due to the invariance propeties of the least squares estimator $\hat{a}_1$ discussed in Section 2 of Andrews (1993), we may consider the case where $m_1 = 0$ and $\sigma^2 = 1$. Secondly, the AR(1) process Y_t for $t = 1, ..., T$ can be inverted and be expressed in terms of residuals. In matrix format, this expression can be presented as follows

$$Y = \Psi\epsilon,$$

with

$$\Psi = \begin{pmatrix} \delta & 0 & 0 & \cdots & 0 & 0 \\ \delta\hat{a}_1 & 1 & 0 & \cdots & 0 & 0 \\ \delta\hat{a}_1^2 & \hat{a}_1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \delta\hat{a}_1^T & \hat{a}_1^{T-1} & \hat{a}_1^{T-2} & \cdots & \hat{a}_1 & 1 \end{pmatrix},$$

where the vector $Y = (y_0, ..., y_T)'$ is expressed in terms of the residuals $\epsilon = (\varepsilon_0, ..., \varepsilon_T)'$, with $\epsilon \sim N(0, \Sigma_{T+1})$, such that 0 is $(T+1) \times 1$ zero vector and Σ is a $T+1$ identity matrix. With such representation, we can now express probability (3.79) as

$$P[\hat{a}_1 \leq c] = P[\epsilon' W_{\hat{a}_1,c}\epsilon \leq 0] = q_{p_j}(\hat{a}_1), \tag{3.80}$$

where $W_{\hat{a},c}$ is a symmetric weight matrix such that

$$W = \Psi' \left[\Delta_0'(I-P)\Delta_T/2 + \Delta_T'(I-P)\Delta_0/2 - c\Delta_T'(I-P)\Delta_T\right]\Psi. \tag{3.81}$$

with $\Delta_0 = (0\vdots I) \in \mathbb{R}^{T\times(T+1)}$, $\Delta_T = (I\vdots 0) \in \mathbb{R}^{T\times(T+1)}$, and I is a T dimensional identity matrix, and finally $P = X(X'X)^{-1}X'$ with $X = (\mathbb{1}_T, \mathbb{T}) \in \mathbb{R}^{T\times 2}$, where $\mathbb{1}_T = (1, ..., 1)' \in \mathbb{R}^T$ and $\mathbb{T} = (1, 2, 3, ..., T)' \in \mathbb{R}^T$. Considering that $\epsilon' W_{\hat{a},c}$ is a quadratic form in normal standard variates, the probabilities $P[\hat{a}_1 \leq c]$ can be computed using the Imhof (1961) approximation in an iterative process involving different values of c, until the desired quantile is obtained.

Other approaches suggested by Andrews (1993) for calculating the quantile of $\hat{a}_1$ are based on simulations and numerical complex integration. These methods are much more suitable when the sample size is greater than $T > 100$, as the probabilities are calculated very slowly by the Imhof (1961) approximation in large sample sizes. ■

The Imhof (1961) algorithm. The Imhof (1961) algorithm concerns evaluating expressions of the form (3.80) - i.e.

$$P[\epsilon' W_{\hat{a}_1,c}\epsilon \leq 0]$$

where as before ϵ is a $(T+1)\times 1$ vector of residuals which are normally distributed with mean 0 and variance Σ_{T+1}, and $W_{\hat{a}_1,c}$ is a $(T+1)\times(T+1)$ symmetric weight matrix defined by (3.81). Let $Q(\epsilon) = \epsilon' W_{\hat{a}_1,c}\epsilon$, we know from Scheffé (1959) that if Σ is non-singular, by using a non-singular linear transformation, $Q(\epsilon)$ can be expressed as

$$Q(\epsilon) = \sum_{j=0}^{T} \lambda_j \varepsilon_j^2 \tag{3.82}$$

or more generally

$$Q(\epsilon) = \sum_{j=0}^{T} \lambda_j \chi^2(m_j, \nu_j), \tag{3.83}$$

where $\lambda_0, ..., \lambda_T$s are the non-zero characteristic roots of $W\Sigma$, m_j is the order of multiplicity of the λ_js and ν_j is the non-centrality parameter. Finally, $\chi^2(m_j, \nu_j)$ are independent χ^2 variables, with m_j degree of freedom. The variable $\chi^2(m_j, \nu_j) = (\varepsilon_0 + \nu)^2 + \sum_{i=1}^{h} \varepsilon_i$, where $\varepsilon_0, ..., \varepsilon_h$ are independent unit normal variates. Therefore, evaluating (3.80) is equivalent to evaluating

$$P\left[\sum_{j=0}^{T} \lambda_j \chi^2(m_j, \nu_j) \leq 0\right], \tag{3.84}$$

where the linear combination of the non-central χ^2 variables has characteristic function

$$\phi(u) = \prod_{j=0}^{T} (1 - 2iu\lambda_j)^{-m_j/2} \exp\left(\sum_{j=0}^{T} \frac{ium_j\lambda_j\nu_j}{1 - 2iu\lambda_j}\right), \tag{3.85}$$

where given the characteristic function $\phi(u)$, the CDF (3.84) can be evaluated using the Fourier-inversion formula [see Gil-Pelaez (1951)] as follows

$$P\left[\sum_{j=0}^{T}\lambda_j\chi^2(m_j,\nu_j)\leq 0\right]=F(0)=\frac{1}{2}-\frac{1}{\pi}\int_0^{\infty}\frac{\mathrm{Im}\{\exp(-iu0)\phi(u)\}}{u}du \\ =\frac{1}{2}-\frac{1}{\pi}\int_0^{\infty}\frac{\mathrm{Im}\{\phi(u)\}}{u}du, \tag{3.86}$$

where $\mathrm{Im}\{z\}$ denotes the imaginary part of the complex number z. Expression (3.86) is evaluated using the *CompQuadForm* package, which includes a translation of Koerts and Abrahamse (1969) FORTRAN code in the statistical programming language R. The algorithm evaluates expression (3.86) involves two steps: in the first step, the integral must be truncated by replacing the upper infinite limit with a suitable replacement; and in the second step, the (now) definite integral is evaluated using the Simpson's rule. The Simpson's rule requires splitting the grid into n even intervals, for which a discretization error tolerance level must be specified, while for the truncation, a truncation error tolerance level should be specified. ■

Asymptotic validity of the VAR sieve bootstrap.

(a) To satisfy condition (A) of Meyer and Kreiss (2015), the residuals $\{\epsilon_t : t\in\mathbb{Z}\}$ are assumed to be a strictly stationary ergodic process, such that $\mathbb{E}(\epsilon_t)=0$ and $\mathbb{E}(\epsilon_t\epsilon_t')=\Sigma$, where Σ is a symmetric and positive definite matrix, and $\mathbb{E}(\|\epsilon_t\|^8)<\infty$. Furthermore, we assume that the spectral density function $P(.)$ of Z_t is bounded -i.e. the eigen values of the spectral density matrix Z_t are uniformly bounded away from zero for all frequencies $(-\pi,\pi]$. Under additional mild assumptions discussed in Section 3.3.1, it can be shown that Z_t is invertible and can be expressed as an infinite autoregressive process.

(b) Let $p\in\mathbb{N}$, such that the order p depends on the sample size -i.e. $p=p(T)$, and let $\hat{\Phi}_1(p),...,\hat{\Phi}_p(p)$ be the Yule-Walker estimators of $\Phi_1(p),...,\Phi_p(p)$. We assume that the convergence rate of Yule-Walker estimators towards the finite sample coefficients is $p(T)^2\sum_{j=1}^{p(T)}\|\hat{\Phi}(p(T))-\Phi(p(T))\|=O_p(1)$ as $T\to\infty$.

(c) Finally, condition (C) of Meyer and Kreiss (2015), requires the functional form the functional form of the Granger causality measure to be of the form

$$C(X \to Y) = f\left(\frac{1}{T-p+1}\sum_{t=1}^{T-p+1} g(Z_t, ..., Z_{T+p-1})\right)$$

for some $p \in \{1, ..., T\}$ and functions $g : \mathbb{R}^{pq} \to \mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}$, where $d \geq 1$, with function f fulfilling smoothness assumptions.

Evidently, assumption (a) is necessary to ensure that a one-sided representation of the underlying process as a VAR(∞) exists, and assumption (b) which concerns the convergence rate of the estimated parameters to the underlying autoregressive coefficients is needed to obtain asymptotic validity of the VAR sieve bootstrap. These assumptions have already been fulfilled and discussed earlier in Section 3.3.1. Moreover, the functional form of the estimator of the Granger causality measure is clearly satisfied and the Lipschitz condition is a consequence of the proof of Proposition 5. ■

Bibliography

Abdalla, I. S. and Murinde, V. (1997). Exchange rate and stock price interactions in emerging financial markets: evidence on india, korea, pakistan and the philippines. *Applied financial economics*, 7(1):25–35.

Aggarwal, R. (1998). Exchange rates and stock prices: A study of u.s. capital markets under floating exchange rates. *Akron Business and Economics Review*, 22(2):7–12.

Andrews, D. W. (1987). Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica: Journal of the Econometric Society*, pages 1465–1471.

Andrews, D. W. (1992). Generic uniform convergence. *Econometric theory*, 8(2):241–257.

Andrews, D. W. (1993). Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica: Journal of the Econometric Society*, pages 139–165.

Andrews, D. W. and Chen, H.-Y. (1994). Approximately median-unbiased estimation of autoregressive models. *Journal of Business & Economic Statistics*, 12(2):187–204.

Bahadur, R. R. and Savage, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *The Annals of Mathematical Statistics*, 27(4):1115–1122.

Bahmani-Oskooee, M. and Sohrabian, A. (1992). Stock prices and the effective exchange rate of the dollar. *Applied economics*, 24(4):459–464.

Banerjee, A., Dolado, J. J., Galbraith, J. W., Hendry, D., et al. (1993). Co-integration, error correction, and the econometric analysis of non-stationary data. *OUP Catalogue*.

Berg, D. (2009). Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance*, 15(7-8):675–701.

Bielecki, T. R., Jakubowski, J., Vidozzi, A., and Vidozzi, L. (2008). Study of dependence for some stochastic processes. *Stochastic analysis and applications*, 26(4):903–924.

Boldin, M. V., Simonova, G. I., and Tyurin, I. N. (1997). *Sign-based methods in linear statistical models*, volume 162. American Mathematical Soc.

Bouezmarni, T., Rombouts, J. V., and Taamouti, A. (2012). Nonparametric copula-based test for conditional independence with applications to granger causality. *Journal of Business & Economic Statistics*, 30(2):275–287.

Bühlmann, P. et al. (1997). Sieve bootstrap for time series. *Bernoulli*, 3(2):123–148.

Bunn, O., Staal, A., Zhuang, J., Lazanas, A., Ural, C., and Shiller, R. (2014). Es-cape-ing from overvalued sectors: Sector selection based on the cyclically adjusted price-earnings (cape) ratio. *The Journal of Portfolio Management*, 41(1):16–33.

Cakan, E. and Ejara, D. D. (2013). On the relationship between exchange rates and stock prices: Evidence from emerging markets. *International Research Journal of Finance and Economics*, (111).

Campbell, B. and Dufour, J.-M. (1991). Over-rejections in rational expectations models: A non-parametric approach to the mankiw-shapiro problem. *Economics Letters*, 35(3):285–290.

Campbell, B. and Dufour, J.-M. (1995). Exact nonparametric orthogonality and random walk tests. *The Review of Economics and Statistics*, 77(1):1–16.

Campbell, B. and Dufour, J.-M. (1997). Exact nonparametric tests of orthogonality and random walk in the presence of a drift parameter. *International Economic Review*, pages 151–173.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*, 1(3):195–228.

Campbell, J. Y. and Shiller, R. J. (2001). Valuation ratios and the long-run stock market outlook: An update. Technical report, National bureau of economic research.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.

Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of financial economics*, 81(1):27–60.

Candelon, B., Joëts, M., and Tokpavi, S. (2013). Testing for granger causality in distribution tails: An application to oil markets integration. *Economic Modelling*, 31:276–285.

Cavanagh, C. L., Elliott, G., and Stock, J. H. (1995). Inference in models with nearly integrated regressors. *Econometric theory*, 11(5):1131–1147.

Coudin, E. and Dufour, J.-M. (2004). Hodges-lehmann sign-based estimators and generalized confidence distributions in linear median regressions with heterogeneous dependent errors. *CIREQ/CIRANO, Department of Economics, McGill University.*

Coudin, E. and Dufour, J.-M. (2009). Finite-sample distribution-free inference in linear median regressions under heteroscedasticity and non-linear dependence of unknown form. *The Econometrics Journal*, 12:S19–S49.

Czado, C., Schepsmeier, U., and Min, A. (2012). Maximum likelihood estimation of mixed c-vines with application to exchange rates. *Statistical Modelling*, 12(3):229–255.

Denuit, M. and Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, 93(1):40–57.

Diks, C. and Panchenko, V. (2006). A new statistic and practical guidelines for nonparametric granger causality testing. *Journal of Economic Dynamics and Control*, 30(9-10):1647–1669.

Dissmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69.

Dufour, J., Jouneau, F., and Torrès, O. (2008). On (non)-testability: Applications to linear and nonlinear semiparametric and nonparametric regression models. Technical report, Technical report, Department of Economics and CIREQ, McGill University, and ....

Dufour, J.-M. (1990). Exact tests and confidence sets in linear regressions with autocorrelated errors. *Econometrica: Journal of the Econometric Society*, pages 475–494.

Dufour, J.-M. and Iglesias, E. M. (2008). Finite sample and optimal adaptive inference in possibly nonstationary general volatility models with gaussian or heavy-tailed errors. Technical report, Working paper, University of Montreal.

Dufour, J.-M. and Jasiak, J. (2001). Finite sample limited information inference methods for structural equations and models with generated regressors. *International Economic Review*, 42(3):815–844.

Dufour, J.-M. and Taamouti, A. (2010a). Exact optimal inference in regression models under heteroskedasticity and non-normality of unknown form. *Computational Statistics & Data Analysis*, 54(11):2532–2553.

Dufour, J.-M. and Taamouti, A. (2010b). Short and long run causality measures: Theory and inference. *Journal of Econometrics*, 154(1):42–58.

Dufour, J.-M. and Taamouti, M. (2003). Point-optimal instruments and generalized anderson-rubin procedures for nonlinear models. Technical report, Technical report, CRDE, Université de Montréal.

Dufour, J.-M. and Torrès, O. (1998). Union-intersection and sample-split methods in econometrics with applications to ma and sure models. *Statistics Textbooks and Monographs*, 155:465–506.

Durante, F. and Sempi, C. (2010). Copula theory: an introduction. In *Copula theory and its applications*, pages 3–31. Springer.

Elliott, G., Müller, U. K., and Watson, M. W. (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica*, 83(2):771–811.

Elliott, G. and Stock, J. H. (1994). Inference in time series regression when the order of integration of a regressor is unknown. *Econometric theory*, 10(3-4):672–700.

Evans, G. and Savin, N. (1981). The calculation of the limiting distribution of the least squares estimator of the parameter in a random walk model. *The Annals of Statistics*, pages 1114–1118.

Evans, G. and Savin, N. (1984). Testing for unit roots: 2. *Econometrica (pre-1986)*, 52(5):1241.

Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of financial economics*, 22(1):3–25.

Furno, M. (2014). Predictions in quantile regressions. *Open Journal of Statistics*, 4(07):504.

Genest, C., Quessy, J.-F., and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33(2):337–366.

Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics*, 44(2):199–213.

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. *Journal of the American statistical association*, 77(378):304–313.

Geweke, J. (1984). Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association*, 79(388):907–915.

Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika*, 38(3-4):481–482.

Goffe, W. L., Ferrier, G. D., and Rogers, J. (1994). Global optimization of statistical functions with simulated annealing. *Journal of econometrics*, 60(1-2):65–99.

Golez, B. and Koudijs, P. (2018). Four centuries of return predictability. *Journal of Financial Economics*, 127(2):248–263.

Gourieroux, C., Monfort, A., and Renault, E. (1987). Kullback causality measures. *Annales d'Economie et de Statistique*, pages 369–410.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438.

Haff, I. H. (2012). Comparison of estimators for pair-copula constructions. *Journal of Multivariate Analysis*, 110:91–105.

Haff, I. H., Aas, K., and Frigessi, A. (2010). On the simplified pair-copula construction—simply useful or too simplistic? *Journal of Multivariate Analysis*, 101(5):1296–1310.

Hatemi-J, A. and Irandoust, M. (2002). On the causality between exchange rates and stock prices: A note. *Bulletin of Economic Research*, 54(2):197–203.

Heinrich, L. (1982). Factorization of the characteristic function of a sum of dependent random variables. *Lithuanian Mathematical Journal*, 22(1):92–100.

Hiemstra, C. and Jones, J. D. (1994). Testing for linear and nonlinear granger causality in the stock price-volume relation. *The Journal of Finance*, 49(5):1639–1664.

Hong, Y., Liu, Y., and Wang, S. (2009). Granger causality in risk and detection of extreme risk spillover between financial markets. *Journal of Econometrics*, 150(2):271–287.

Imhof, J.-P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3/4):419–426.

Jansson, M. and Moreira, M. J. (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica*, 74(3):681–714.

Jeong, K., Härdle, W. K., and Song, S. (2012). A consistent nonparametric test for causality in quantile. *Econometric Theory*, 28(4):861–887.

Jivraj, F. and Shiller, R. J. (2017). The many colours of cape.

Joe, H. (1996). Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters. *Lecture Notes-Monograph Series*, pages 120–141.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Joe, H. (2014). *Dependence modeling with copulas*. Chapman and Hall/CRC.

King, M. L. (1987). Towards a theory of point optimal testing. *Econometric Reviews*, 6(2):169–218.

Koerts, J. and Abrahamse, A. P. J. (1969). *On the theory and application of the general linear model*. Rotterdam University Press.

Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies*, 28(5):1506–1553.

Krueger, A. O. (1983). *Exchange-rate determination*. Cambridge University Press.

Lee, T.-H. and Yang, W. (2012). Money–income granger-causality in quantiles. In *30th Anniversary Edition*, pages 385–409. Emerald Group Publishing Limited.

Lehmann, E. L. and Romano, J. P. (2006). *Testing statistical hypotheses*. Springer Science & Business Media.

Lehmann, E. L., Stein, C., et al. (1949). On the theory of some non-parametric hypotheses. *The Annals of Mathematical Statistics*, 20(1):28–45.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–235.

Li, J. and Wong, W. K. (2010). Two-dimensional toxic dose and multivariate logistic regression, with application to decompression sickness. *Biostatistics*, 12(1):143–155.

Luger, R. (2003). Exact non-parametric tests for a random walk with unknown drift under conditional heteroscedasticity. *Journal of Econometrics*, 115(2):259–276.

Magdalinos, T. and Phillips, P. C. (2009). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory*, 25(2):482–526.

Mankiw, N. G. and Shapiro, M. D. (1986). Do we reject too often?: Small sample properties of tests of rational expectations models. *Economics Letters*, 20(2):139–145.

Maronna, R. A., Martin, R. D., Yohai, V. J., and Salibián-Barrera, M. (2019). *Robust statistics: theory and methods (with R)*. John Wiley & Sons.

McCloskey, D. N., Ziliak, S. T., et al. (1996). The standard error of regressions. *Journal of economic literature*, 34(1):97–114.

Meyer, M. and Kreiss, J.-P. (2015). On the vector autoregressive sieve bootstrap. *Journal of Time Series Analysis*, 36(3):377–397.

Muhammad, N., Rasheed, A., and Husain, F. (2002). Stock prices and exchange rates: Are they related? evidence from south asian countries [with comments]. *The Pakistan Development Review*, pages 535–550.

Nelson, C. R. and Kim, M. J. (1993). Predictable stock returns: The role of small sample bias. *The Journal of Finance*, 48(2):641–661.

Nikoloulopoulos, A. K. and Karlis, D. (2008). Multivariate logit copula model with an application to dental data. *Statistics in Medicine*, 27(30):6393–6406.

Nikoloulopoulos, A. K. and Karlis, D. (2009). Finite normal mixture copulas for multivariate discrete data modeling. *Journal of Statistical Planning and Inference*, 139(11):3878–3890.

O'Gorman, T. W. (2004). *Applied adaptive statistical methods: tests of significance and confidence intervals*, volume 12. Siam.

Oh, D. H. and Patton, A. J. (2016). High-dimensional copula-based distributions with mixed frequency data. *Journal of Econometrics*, 193(2):349–366.

Olugbenga, A. A. (2012). Exchange rate volatility and stock market behaviour: the nigerian experience. *European Journal of Business and Management*, 4(5).

Panagiotelis, A., Czado, C., and Joe, H. (2012). Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association*, 107(499):1063–1072.

Phillips, P. C. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica*, 82(3):1177–1195.

Phillips, P. C. and Lee, J. H. (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics*, 177(2):250–264.

Phillips, P. C. and Lee, J. H. (2016). Robust econometric inference with mixed integrated and mildly explosive regressors. *Journal of Econometrics*, 192(2):433–450.

Phillips, P. C., Magdalinos, T., et al. (2009). Econometric inference in the vicinity of unity. *Singapore Management University, CoFie Working Paper*, 7.

Phillips, P. C. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346.

Polasek, W. (1994). Temporal causality measures based on aic. In *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: an Informational Approach*, pages 159–168. Springer.

Polasek, W. (2000). Bayesian causality measures for multiple arch models using marginal likelihoods.

Pratt, J. W. and Gibbons, J. D. (2012). *Concepts of nonparametric theory*. Springer Science & Business Media.

Rozeff, M. S. (1984). Dividend yields are equity risk premiums. *Journal of Portfolio management*, pages 68–75.

Rudebusch, G. D. (1992). Trends and random walks in macroeconomic time series: A re-examination. *International Economic Review*, pages 661–680.

Scheffé, H. (1959). The analysis of variance john wiley & sons. *Inc., New York*, pages 351–358.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

Soenen, L. and Aggarwal, R. (1989). Financial prices as determinants of changes in currency values. In *25th Annual Meetings of Eastern Finance Association, Philadelphia*, volume 2, pages 2–2.

Soenen, L. A. and Hennigar, E. S. (1988). An analysis of exchange-rates and stock-prices-the united-states experience between 1980 and 1986. *Akron Business and Economic Review*, 19(4):7–16.

Song, P. X.-K., Li, M., and Yuan, Y. (2009). Joint regression analysis of correlated data using gaussian copulas. *Biometrics*, 65(1):60–68.

Song, X. and Taamouti, A. (2018). Measuring nonlinear granger causality in mean. *Journal of Business & Economic Statistics*, 36(2):321–333.

Stambaugh, R. F. (1985). *Bias in regressions with lagged stochastic regressors*. Center for Research in Security Prices, Graduate School of Business ....

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3):375–421.

Stock, J. H. (1991). Confidence intervals for the largest autoregressive root in us macroeconomic time series. *Journal of monetary economics*, 28(3):435–459.

Stoeber, J., Joe, H., and Czado, C. (2013). Simplified pair copula constructions—limitations and extensions. *Journal of Multivariate Analysis*, 119:101–118.

Su, L. and White, H. (2008). A nonparametric hellinger metric test for conditional independence. *Econometric Theory*, 24(4):829–864.

Su, L. and White, H. (2014). Testing conditional independence via empirical likelihood. *Journal of Econometrics*, 182(1):27–44.

Taamouti, A. (2015). Finite-sample sign-based inference in linear and nonlinear regression models with applications in finance. *L'Actualité économique*, 91(1-2):89–113.

Taamouti, A., Bouezmarni, T., and El Ghouch, A. (2014). Nonparametric estimation and inference for conditional density based granger causality measures. *Journal of Econometrics*, 180(2):251–264.

Toda, H. Y. and Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of econometrics*, 66(1-2):225–250.

Torous, W., Valkanov, R., and Yan, S. (2004). On predicting stock returns with nearly integrated explanatory variables. *The Journal of Business*, 77(4):937–966.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *econometrica*, 48(4):817–838.

White, H. (2014). *Asymptotic theory for econometricians.* Academic press.

Wiener, N. (1956). Nonlinear prediction and dynamics. In *Third Berkeley Symposium on Mathematical Statistics and Probability*, pages 247–252.

Zeren, F. and Koç, M. (2016). Time varying causality between stock market and exchange rate: evidence from turkey, japan and england. *Economic research-Ekonomska istraživanja*, 29(1):696–705.